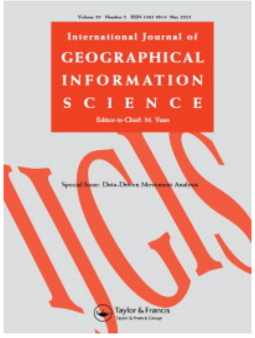


Taylor & Francis
Taylor & Francis Group

International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tgis20

Decoding wayfinding: analyzing wayfinding processes in the outdoor environment

Negar Alinaghi, Ioannis Giannopoulos, Markus Kattenbeck & Martin Raubal

To cite this article: Negar Alinaghi, Ioannis Giannopoulos, Markus Kattenbeck & Martin Raubal (18 Mar 2025): Decoding wayfinding: analyzing wayfinding processes in the outdoor environment, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2025.2473599

To link to this article: https://doi.org/10.1080/13658816.2025.2473599

Published online: 18 Mar 2025.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2025.2473599

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Decoding wayfinding: analyzing wayfinding processes in the outdoor environment

Negar Alinaghi[a], Ioannis Giannopoulos[a], Markus Kattenbeck[a] and Martin Raubal[b]

[a]Department of Geodesy and Geoinformation, TU Wien, Vienna, Austria; [b]Institute of Cartography and Geoinformation, ETH Zurich, Zurich, Switzerland

ABSTRACT

Navigating complex environments is crucial for human life, yet understanding the cognitive processes involved in its wayfinding component remains challenging. One theoretical model that explains these processes is Downs and Stea's four-step model. Our study builds on this model to empirically analyze its steps, focusing particularly on the monitoring step. Machine learning models were trained on gaze behavior and head/body movement data from over 300 routes walked by 56 participants in a real-world outdoor study, predicting three of these wayfinding steps: self-localization, route planning, and goal recognition. Applying this trained model to the respective monitoring segment of the same routes suggests that monitoring includes micro-versions of these three steps, indicating it operates as a recursive process rather than a distinct cognitive step. By bridging theoretical frameworks with empirical evidence, these findings enhance our understanding of spatial cognition and can inform the design of navigational tools and urban spaces.

ARTICLE HISTORY
Received 6 September 2024
Accepted 24 February 2025

KEYWORDS
Wayfinding behavior; cognitive processes; machine learning; eye-tracking; head movement tracking

1. Introduction

Navigating complex environments is an essential part of human life. Wayfinding is at the core of this ability, a multifaceted cognitive process involving orientation, decision-making, and environmental perception (Montello and Sas 2006). Studying these processes offers valuable insights into spatial cognition and perception, enhancing our understanding of how individuals interpret and interact with their surroundings. This understanding can inform the design of navigational tools and urban spaces, improving accessibility and efficiency (Seidel 1982, Arthur and Passini 1992, Raubal 2001).

Wayfinding research is grounded in a rich history of theoretical frameworks, each offering unique perspectives on its cognitive processes. Lynch (1964) was the first to explore the visual quality of cities and the role of environmental cues in navigation.

CONTACT Negar Alinaghi negar.alinaghi@geo.tuwien.ac.at

Later, Downs and Stea (1977) proposed a four-step model of wayfinding processes. Passini (1981) framed wayfinding as spatial problem-solving with three phases. Montello (2001) distinguished wayfinding from locomotion as the two components of navigation and explored spatial cognition's role in wayfinding (Montello and Raubal 2012).

Despite extensive theoretical work, empirical studies validating these frameworks are limited. While several studies (Dong *et al*. 2022, Dalton 2003, Kiefer *et al.* 2014, Yang *et al.* 2020, Alinaghi *et al*. 2022, 2023) have explored individual aspects of wayfinding, comprehensive empirical evaluation of the full wayfinding span and its cognitive processes is lacking. This gap may result from the complexity of cognitive processes and practical measurement challenges. However, empirically studying these processes is important for refining existing theories or developing new ones.

Among the theoretical frameworks, Downs and Stea (1977)'s four-step model stands out for its detailed breakdown of cognitive processes:

> …we can break down the process [of wayfinding] into four sequential and interrelated steps: (1) orientation, (2) the choice of route, (3) keeping on the right track, and (4) the discovery of the objective. (Downs and Stea 1977, p. 124)

These steps—also known as self-localization, route planning, monitoring, and goal recognition—can be behaviorally observed and measured, making them more suitable for empirical investigation than models focused solely on decision-making processes.

However, Downs and Stea's model presents challenges in understanding the interrelatedness and transitions between its sequential steps. The monitoring step, defined as 'keeping on the right track', is particularly vague. While presented as a standalone step, monitoring could actually involve all other steps—continuous awareness of the current location, planning a route, and recognizing the destination. Yet these actions might differ from initial self-localization, initial route planning, and final goal recognition, in that they may only need to be updated and fine-tuned once realized for the first time. Thus, monitoring could be seen as involving micro-versions of these three steps: To fine-tune self-localization, route planning, and goal recognition, one might recursively set intermediate goals, continually self-localize, plan a route, and confirm the achievement of these micro-goals until reaching the final destination.

This study examines two interpretations of the monitoring step: The first is as a standalone step, as defined by Downs and Stea. The second is as a recursive call of the other three steps (i.e. self-localization, route planning, and goal recognition). The required analysis for this is structured into two parts: First, we use machine learning (ML) to train a model predicting three of the four steps of wayfinding as defined by Downs and Stea, based on gaze behavior and head/body movements. We refer to these instances used for training as the *macro* versions of these three activities. Second, we use this trained model to decode the monitoring step, examining whether the same learned patterns appear and how these patterns, referred to as the *micro* versions of the three activities, are distributed during monitoring and related to the macro steps.

56 participants navigated over 300 routes in familiar and unfamiliar environments while wearing eye-tracking glasses and a head-mounted IMU (Inertial Measurement Unit) sensor. Tasks reflected the four-step model: determining the current location,

drawing a route to a given destination, following it, and verifying destination recognition. The phase of following the planned route until verifying the destination was considered the monitoring step. The classifier predicting the three macro steps achieved 87.8% accuracy and a kappa value of $\kappa = 81.9\%$ on unseen data. This model was then used to infer the micro steps during monitoring using sequence analysis techniques.

Results suggest the presence of micro-versions of the three steps during monitoring, supporting the idea that monitoring could be seen as a recursive call of the same three functions on a micro-scale. By bridging theoretical frameworks with empirical evidence, our study advances wayfinding research and offers insights into human navigation behavior for practical applications in various domains.

2. Related work

Subsection 2.1 introduces key theoretical models of wayfinding with a focus on Downs and Stea's model. Subsection 2.2 then reviews empirical studies linking various aspects of wayfinding to these theories. This highlights the need for a comprehensive empirical study to decode the complex cognitive processes involved in wayfinding.

2.1. Theoretical models of wayfinding

Montello (2001) defines wayfinding and locomotion as two components of navigation. Wayfinding, the cognitive process of orienting oneself in physical spaces, has been studied in various disciplines. Lynch (1964) initiated this research by exploring how urban landscapes shape mental images that are essential for successful navigation. His work highlighted the role of mental representations of the space—a fusion of sensory inputs and memories—in guiding spatial understanding and actions, thereby laying the foundation for understanding how environmental features influence wayfinding strategies. As mentioned in Section 1, Downs and Stea (1977) introduced a cognitive model of wayfinding that is framed as a four-step cycle (in this order): self-localization, route planning, monitoring, and goal recognition. These steps are defined as follows:

> '*We must know where we are* in relation to some selected places on the Earth's surface'. (p. 124) … 'The *choice of a route* requires that a person make a cognitive connection between his current location and that of the desired destination. To be useful, this connection must be converted into a plan of actions … ' (p. 130) … 'The third step in the process, *keeping on the right track*, monitors the execution of the route plan … keeping on the right track is achieved by keeping our cognitive map tied to the perceived environment, by making and taking appropriate actions at each decision point'. (p. 133) … '*You must recognize that you have got to where you are going*. Again, this recognition depends upon linking a cognitive map with the perceived environment, a linkage that usually brings profound relief'. (p. 135)

Downs and Stea provide a detailed breakdown of the process into sequential and interrelated steps. Although some aspects, such as the definition of monitoring and the interrelationships between the steps, are not clear enough, the model is promising for empirical testing since the defined steps are behaviorally observable and measurable.

Passini (1981) conceptualized wayfinding as a dynamic problem-solving process, highlighting individualized strategies in different contexts. He identified 'three distinct but not necessarily chronological phases': processing information, making decisions based on the processed information, and taking actions based on the decisions. In a review article, Jamshidi and Pati (2021) discussed various problem-solving models related to wayfinding, including the Test–Operate–Test–Exit (TOTE) model (Miller *et al.* 2017), self-regulation theories (e.g. Carver and Scheier (2001)), and progress monitoring theory (MacGregor *et al.* 2001). These models share a feedback-driven cycle in which individuals test, adjust, and repeat until reaching their goal. Although these models cover the full span of wayfinding, they focus on decision-making, making the associated actions difficult to observe empirically.

Several other models address wayfinding by focusing on the cognitive functions required for successful wayfinding rather than defining the sequence of steps. For example, Chown *et al.* (1995) integrated cognitive mapping, emphasizing landmark recognition, path selection, direction choice, and spatial abstraction for successful wayfinding. Golledge (1999) categorized wayfinding into exploratory navigation, travel to familiar destinations, and travel to unfamiliar destinations, outlining the cognitive functions needed for each. Golledge (2003) expanded this by categorizing related tasks like path integration, piloting, and chunking, showing how spatial knowledge and perceptual cues guide navigating un/familiar environments. Wiener *et al.* (2009) offered an extended taxonomy, categorizing tasks based on external constraints and the wayfinder's level of spatial knowledge. Montello and Raubal (2012) explored spatial cognition's role in key wayfinding functions like route planning, landmark recognition, and distance estimation, explaining how these functions are essential for successful wayfinding.

These perspectives advance our understanding of wayfinding as a complex interplay of cognitive processes and environmental cues. Despite different approaches, some cognitive processes or tasks, e.g. route planning and landmark recognition are consistently identified by different researchers. We selected the model by Downs and Stea for its detailed categorization of steps that span the entire wayfinding task and are defined in a way that can be measured and analyzed empirically, at least to a certain extent.

2.2. Empirical studies for understanding wayfinding processes

Recent empirical studies from various disciplines focusing on key wayfinding behaviors, e.g. self-localization and route planning, have advanced our understanding of wayfinding and spatial cognition. Among the various sensors used, eye-tracking has become a valuable method for studying these behaviors, as it reveals how gaze behavior reflects underlying cognitive processes. Since the work of Yarbus (1967), gaze behavior has captured significant attention from researchers, with studies showing that different cognitive tasks produce distinct gaze patterns (Just and C 1976). In particular, research on eye movements in spatial decision-making (see Kiefer *et al.* (2017) for an overview) highlighted the importance of eye-tracking in understanding how individuals process spatial information. For example, Kiefer *et al.* (2014) used mobile eye-tracking technology to explore self-localization in urban environments, and found that effective self-localization involves increased visual attention to map symbols and shifts between these symbols and real-world landmarks. Meilinger *et al.* (2007) compared different schematic maps for self-localization in complex buildings and reported similar performance across various map types. Schmid *et al.* (2010) developed a map design that integrates local and global orientation cues, improving the speed and accuracy of self-localization. Brügger *et al*. (2018) used eye-tracking to investigate how navigational cues influence spatial orientation, observing fewer forward and backward glances with no significant effect on sideways glances.

From a neuroscience perspective, Epstein and Vass (2014) showed how humans and animals use landmarks for navigation, linking landmark recognition, localization, and spatial knowledge to specific regions of the brain. More recently, Peer *et al.* (2023) showed through the analysis of brain activities that the accuracy and variability of cognitive maps depend on the structure of the environment, supporting the theory of Lynch (1964). Similarly, Hartley *et al*. (2003) used fMRI to show that successful wayfinders activate the caudate nucleus during route-following, aligning with research showing that wayfinding relies on both caudate (response-based) and hippocampal (place-based) representations.

In a closely related paper, (Alinaghi and Giannopoulos 2024) explored the visual attention of familiar and unfamiliar wayfinders during the different steps of wayfinding as defined by the Downs and Stea (1977) model. The authors found that distinct gaze and visual attention patterns emerge at different steps of wayfinding, with some patterns significantly influenced by the individual's spatial familiarity. Building on Passini's wayfinding model (Passini 1981), Spiers and Maguire (2008) conducted an empirical study using retrospective verbal reports and eye-tracking data from 20 participants performing wayfinding tasks in a virtual reality simulation of London, UK. Participants described their thoughts during wayfinding, and the authors classified them and computed their frequency, duration, and temporal order. The findings reveal a wider range of thoughts than the theory suggests but confirm that 'route planning' and 'action planning', as defined by the model, are the most frequent and central processes in wayfinding cognition.

While these studies offer important empirical insights into wayfinding, they mainly focus on specific aspects of this complex cognitive process. Therefore, there is still a gap in the empirical study of the entire process of wayfinding by observing the behaviors and actions of wayfinders. This is not only crucial for improving related technologies or applications, but also for creating theoretical frameworks or refining existing ones.

3. Experimental design and procedure

To empirically study the wayfinding processes defined by Downs and Stea, we designed an outdoor real-world human-subject study[1], conducted in Vienna, Austria. Data[2] was collected in two steps—an online step and an on-site step—between

August 2021 and June 2023 due to challenges, e.g. weather and daylight variability, and COVID-related interruptions. The following subsections detail these two steps.

3.1. Step 1: Online data collection

Given the significant impact of personal attributes on wayfinding, we included an online registration step to collect personal data, which took roughly 15-20 minutes to complete. To ensure a comprehensive understanding of each participant's background and spatial abilities, we collected demographic data, participants' level of spatial familiarity with different regions in Vienna, preferences for navigational aids, responses to the SBSOD (Santa Barbara Sense of Direction) questionnaire (Hegarty 2002), and Short 15-item Big Five Inventory test (Lang *et al.* 2011). The spatial familiarity data was then used to customize the subsequent on-site data collection. The following subsections provide details on the participants and the familiarity assessment.

3.1.1. Participants

Participants were recruited through a snowball method, where colleagues and friends who had lived in Vienna for at least three months were asked to refer additional participants. Residency was important to ensure familiarity with, at least, parts of the city. A total of 84 individuals completed the online part of the experiment, but only 67 attended the on-site step[3]. The rest were non-responsive and three were excluded due to wearing correction glasses with a prescription above ± 3.5, which would affect eye-tracking accuracy. Of the 67, we experienced data loss for 11 participants due to sensor malfunction (five experiments were terminated early, and for six the data loss was detected after data collection). The final sample included 56 participants ($mean_{age} = 31.16$ years, $std_{age} = 5.93$) with 22 females and 34 males.

3.1.2. Familiarity assessment

After registration, participants were directed to a web map and asked to rate their familiarity with different regions in Vienna (presented as a hexagonal tessellation) on a 5-point scale, where 1 indicated completely unfamiliar and 5 indicated completely familiar. To ensure a comprehensive familiarity assessment, participants were required to rate the entire city. The inner city was covered by 1.00 km^2 cells by default, and the outer city by 7.00 km^2 cells (which could be divided by the user into seven 1.00 km^2 cells), assuming that the inner city, being more touristic, would likely have common familiar cells for most participants.

3.2. Step 2: On-site data collection

After the online step, we used the collected familiarity data to select a familiar and an unfamiliar study area for each participant. To ensure consistent urban design and environmental characteristics, we selected both areas from the inner city's 1.00 km^2 hexagonal grids. These areas have similar urban characteristics, considering building age and architectural design (Reimer *et al.* 2022). For each participant, the cells with the highest and lowest familiarity scores were selected from these areas.

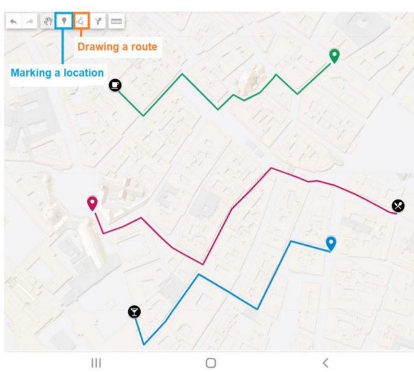
Within each cell, multiple OD (Origin-Destination) pairs were selected, with a walking distance of 7–10 minutes and at least one turn, due to the behavioral differences observed at turning versus non-turning junctions (i.e. intersections where participants continued straight ahead) (Alinaghi *et al*. 2022). Local landmarks, such as shops, restaurants, and cafes, served as destinations. It was planned that each participant would walk three different routes in both familiar and unfamiliar areas, resulting in six routes per participant. The entire procedure, including sensor setup and calibration, took roughly 1 hour 45 minutes per participant. Various hardware and software tools were used to collect data for each trial, as detailed in the following subsections.

3.2.1. Hardware and software

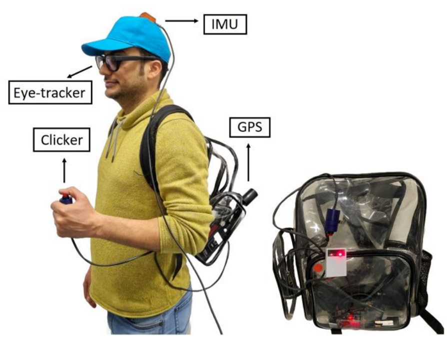
The outdoor data collection used several sensors and hardware (Figure 1(b) shows the setup of these sensors on the participant):

- Eye-Tracker: PupilLabs Invisible glasses (Tonsen *et al*. 2020) were used to record eye movements at a frequency of 200Hz running on an Android phone.
- IMU Sensor: the xSens MTi-300 IMU (Xsens Technologies 2020), positioned on top of the head to be aligned with the forward-facing view, was used to capture data such as velocity, acceleration, rotation, and magnetic field at frequencies from 100 to 1000Hz. This sensor was connected to a Windows-based laptop.
- GPS: the PPM 10-xx38 GNSS receiver, connected to an Android phone receiving EPOSA ground corrections, recorded location data at 1Hz.
- Clicker: Participants used a custom-built clicker device connected to an LED light on their backpack to indicate recognition of the destination by clicking it.

Two customized Android applications facilitated outdoor data collection:



(a) The participant's app was a customized Google Maps application with specific functionalities designed to facilitate the tasks corresponding to the model by Downs and Stea (1977).



(b) A participant equipped with an eye-tracker, an IMU attached to the cap, a GPS receiver on the backpack, and a handheld clicker connected to an LED light attached to the backpack. The participant has provided written permission for the use of his photo in publications.

Figure 1. (a) Shows the participants' app interface used for self-localization and route planning tasks; and (b) illustrates the sensor setup on the participant during the outdoor data collection.

- Participant's app, shown in Figure 1(a), resembled an adapted version of Google Maps with modified functions. To enable self-localization, both the user's current location and the location search functions were disabled. The app included a line drawing tool for route planning, and while being zoomable, at the beginning, it was displayed in the selected familiarity cell by default. Pre-selected destinations were marked on the map with their names for the unfamiliar condition, and their names and their Google Street View images for the familiar condition. This approach aimed to increase the sense of familiarity in the familiar condition and enable more realistic wayfinding, as it was not certain that participants were familiar with all the selected destinations in the familiar areas. Map visualization was standardized with the default Google base map and the tablet's screen rotation was locked to portrait view in order to control visualization preferences.
- Experimenter's app included five simple trigger buttons to log different events during the wayfinding task (Subsection 3.2.2). These logged events were labeled in more detail later during data cleaning and were synchronized with sensor recordings to serve as ground truth labels for analysis.

3.2.2. Procedure

Participants received instructions via email on where to meet on-site, with half of them randomly assigned to the pre-selected familiar area and the other half to the pre-selected unfamiliar area. This randomization was essential to minimize learning biases, as starting with the same condition could lead to relaxation and learning effects that would affect performance in the second condition. Outdoor trials were structured as follows:

1. Sensor Setup: Sensors were set up and calibrated.
2. Instructions for the Wayfinding Tasks:

    a. *Self Localization:* Participants were asked to find their current location and mark it in the participants' app (i.e. the customized Google Maps app), which was provided on an 11-inch tablet with a 2560×1600 LCD display.
    b. *Orientation:* The experimenter pointed in a direction in the surroundings by hand and asked participants to show on the map where they would go if they walked in that direction.
    c. *Route Planning:* Participants were instructed to draw the route they preferred to take to the destination displayed to them on the map.
    d. *Walk to the Destination:* Participants were asked to walk their planned route while maintaining their normal behavior (as if they were alone) and being focused on the wayfinding task. They were allowed to use the map as many times as they wanted.
    e. *Goal Recognition:* Participants were asked to press the clicker in their hand upon spotting and recognizing the destination while still walking toward it.

To optimize resource-intensive data collection, we increased the number of routes per participant using a cognitive reset strategy to minimize route recollection. After

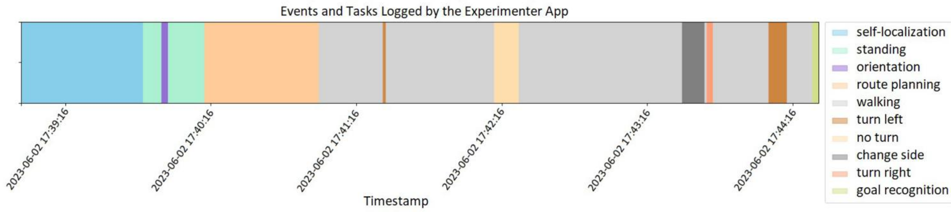


Figure 2. The sequence and duration of recorded events during a sample route. Detailed labels (e.g. "turn left/right/no turn" instead of "junction") are derived during data processing.

reaching a destination, participants read aloud from a text while being escorted to a new starting point (2–3 minutes walk), resetting self-localization by distracting them from observing their surroundings. This reset method was also used in the unfamiliar condition before the first route to avoid familiarity with the meeting point. Each trial included, on average, three routes, with faster participants completing up to four routes and slower ones completing two. The same setup was used for the second condition where the participant was tested at a different location. During each route, the experimenter logged the following events using the experimenter app (Figure 2):

- *Self Localization:* From identifying the current location to marking it on the map.
- *Orientation:* From the moment the direction was shown to the participant (exclusive) until they showed the direction on the map.
- *Route Planning:* From showing the destination to the participant on the map (exclusive) to when they finished drawing the route.
- *Events During Walking:*
  - *Participant Behavior:* Duration of behaviors such as feeling lost (e.g. slowing down, standing, looking around) or making mistakes (e.g. wrong turns, missing the destination). Unexpected events, like interruptions from the experimenter (e.g. sensor failure) or curious pedestrians, were also logged.
  - *Junction:* Duration of passing a junction.
  - *Change Side:* Duration of walking from one side to the other of the street.
  - *Clicker:* From clicking the clicker device to indicate destination recognition to having reached the destination.

4. Data pre-processing

This section describes the main pre-processing techniques used to prepare the data for the ML experiments, including data synchronization, cleaning, and feature engineering.

4.1. Data synchronization

Given the multi-sensor data collection, with each sensor operating on its own clock and a different operating system, careful synchronization was crucial. We followed a three-step process to synchronize the sensors: First, we displayed UTC timestamps on all device screens using Android and Python applications. Screens with timestamps

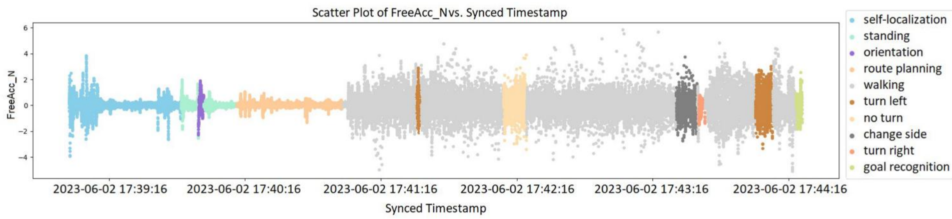


Figure 3. The scatter plot of the free acceleration signal confirms the alignment of IMU and experimenter app data with the master clock (i.e. the eye-tracker time), as evidenced by the clear peak corresponding to the "walking" event.

were captured via the eye tracker's scene camera at the start of gaze recording, serving as a reference point. Second, we calculated time differences between each device and the eye-tracking time (establishing a master time) by extracting frame numbers of video displaying timestamps on other devices. Finally, we applied these time differences to respective recordings to synchronize the timestamps. This process assumes negligible clock drift and constant differences throughout the recording, supported by the short duration of recording and the high-frequency timestamps.

To validate IMU synchronization, we compared its recordings with logged events. IMU records a peak in acceleration signals at the start of walking. Figure 3 shows, as an example from a random route, the free acceleration signal in the Y-axis[4] (FreeAcc-N), and how the start of the peak correctly aligns with the synced label of 'walking'.

4.2. Data cleaning

After synchronizing the data, we checked for errors such as malfunctions or recording interruptions. Recordings with clear issues (e.g. empty or corrupted files, or missing data) were excluded. For more latent errors, we applied the 3σ-rule to IMU and eye-tracker signals, flagged anomalies, visually inspected them, and cross-checked with scene videos. Out-of-distribution values, i.e. any values outside the 3σ threshold, were removed to ensure a clean dataset. Initially, we expected 336 routes (6 routes per participant). However, as not all participants completed 6 routes, and after data cleaning, we were left with 309 valid routes, consisting of 157 familiar and 152 unfamiliar routes.

4.3. Feature extraction

After data preparation, feature extraction was performed for ML experimentation. To transform the recorded signals into tabular data, we used a sliding window approach to segment the data and extract the features for every window.

While factors such as spatial familiarity (Kattenbeck *et al.* 2024), environmental features (Alinaghi *et al.* 2023), and wayfinder characteristics (e.g. demographics, spatial skills, etc. (Montello and Sas 2006)) can influence behaviors during wayfinding, we focused on gaze and head/body movements. Although we collected these data, analyzing them was beyond the scope of this study. To answer our research questions at

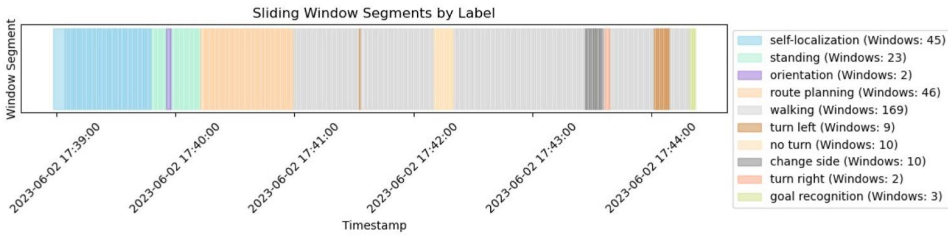


Figure 4. Visualization of the overlapping sliding window applied to a sample route. The legend shows the number of instances generated for each logged event (e.g. 45 instances of self-localization) using the sliding window segmentation.

a general level, we aimed to predict the steps of wayfinding and decode the monitoring step using only gaze and movement features without this contextual information.

4.3.1. Sliding window segmentation

Two sliding window approaches were used based on suggestions in the literature (Dehghani *et al*. 2019): a non-overlapping fixed-size window and an overlapping fixed-size window. The fixed-size approach ensures consistent feature extraction across instances and enhances computational efficiency while maintaining accuracy. For the overlapping method, a step size of one second was used to maximize data.

Window sizes from 2 to 10 seconds were evaluated in terms of classification accuracy and kappa across 18 experiments (9 non-overlapping and 9 overlapping). The optimal segmentation was determined to be a 6-second overlapping window. Therefore, we exclusively report the results from this segmentation. Figure 4 illustrates the application of this segmentation. Each window is referred to as one instance in our dataset. The 6-second overlapping segmentation across 309 routes results in over 90, 000 instances which is a large enough dataset for the ML experiments.

4.3.2. Gaze features

The gaze features extracted for each instance focus on two primary gaze events: fixations and saccades, which are key indicators of cognitive processes, scene comprehension, and visual search (refer to Chapters 13 and 21 of Duchowski (2017)). Fixation-related features include frequency (count of fixations in time intervals), duration (period of relatively still eye position known as dwell time), and dispersion (maximum horizontal and vertical distance covered by gaze positions during fixation, as defined by Salvucci and Goldberg (2000)). Saccadic features include duration (time between two fixations), amplitude (distance traveled by a saccade), and rate (number of saccades per second). To account for the impact of head movements on saccades in mobile eye-tracking, we apply a saccadic correction method from Alinaghi and Giannopoulos (2022) ensuring that the extracted saccadic features are based on corrected data.

Since gaze movement during the experiment partially involved the tablet, we also considered the distribution of fixation points. During route planning, fixations are expected to cluster around the planned route, whereas in self-localization, they tend to be more dispersed. To capture these variations, we computed four key features

Table 1. Calculated gaze features: for each instance, the frequency, mean, minimum, maximum, and variance of various fixation- and saccade-related features, as well as several features encoding the fixation distribution patterns, are computed.

Gaze events	Features	Definition	Count
Fixation	Frequency	# Fixations divided by window size (6 sec)	1
	Mean/min/max/variance – duration	The period when the eye remains relatively still from the first gaze position to the last gaze position of a fixation	4
	Mean/min/max/variance – dispersion	The maximum horizontal and vertical distance covered by the gaze positions in a fixation	4
	Entropy	The entropy of the KDE of fixation points	1
	Intensity	The average number of fixations in the minimum bounding box	1
	Stdd	Standard deviation of distances between fixations	1
	Theta	Angular difference between consecutive fixations	1
Saccade	Frequency	# Saccades divided by window size (6 sec)	1
	Mean/min/max/variance – duration	The time taken to move between two fixations	4
	Mean/min/max/variance – length	The distance traveled by a saccade	4
Total count			22

that encode the distribution of fixation points (Kang *et al.* 2023). We used KDE (Kernel Density Estimation) to estimate the probability density function, which provides a smoothed representation of point distribution in 2D space. We computed the entropy of the KDE, where higher entropy indicates greater dispersion or randomness, and lower entropy suggests more concentrated distributions. In addition, taken from point pattern analysis (Baddeley *et al.* 2016), we calculated intensity and centrography. Intensity measures the average number of event points per unit area and reflects spatial concentration. Two measures were used for centrography: Standard Distance, which is similar to the traditional standard deviation and indicates the dispersion of events around the spatial mean center, and the standard ellipse of dispersion, which assesses the dispersion and orientation in two dimensions. Table 1 summarizes the 22 gaze features extracted for each instance.

4.3.3. IMU features

We used the XSENS MTi-300 IMU sensor to accurately track head and body movements based on a gyro-enhanced Attitude and Heading Reference System (AHRS). It records normal acceleration (velocity changes), free acceleration (gravity-independent), and high-resolution acceleration at 1000 Hz to distinguish motion from stillness and abrupt versus smooth head movements. Gyroscopic measurements indicate rotational velocity captured at 400 Hz and 1000 Hz, and magnetic field strength identifies non-movement anomalies. Velocity changes, orientation, rotation increments, and angles (orientation change and quaternions) are measured at 400 Hz, with Euler angles at 100 Hz tracking precise rotations around longitudinal, lateral, and vertical axes.

Head and body movements are task-dependent and can indicate the task being performed, as shown in human activity recognition research (e.g. Vanrell *et al.* (2018)).

Table 2. Calculated IMU features: statistical measures (mean, minimum, maximum, and variance) from each signal were computed for each instance.

IMU recordings	Features (mean/min/max/ variance)	Definition	Count
Acceleration	*Acceleration X/Y/Z*	The rate of change of velocity along the three axes	12
	free acceleration X/Y/Z	Acceleration in the local frame along the three axes from which the local gravity is deducted	12
	HR acceleration X/Y/Z	Accelerations at a high sampling rate along the three axes	12
3D rate of turn	*gyros X/Y/Z*	The angular rate or rotational velocity measured around the three axes	12
	HR gyros X/Y/Z	The angular rate at a high sampling rate measured around the three axes	12
Magnetic field	*magnetic field X/Y/Z*	The magnetic field strength or intensity along the X, Y, and Z-axis	12
Velocity change	*velocity change X/Y/Z*	Velocity change during a certain interval (this interval is 2.5 ms (400 Hz) by default)	12
Orientation change	*orientation change q0*	The scalar component of the quaternion representing the rotation increment	4
	orientation change q1	The component associated with the rotation about the X-axis in the quaternion representation	4
	orientation change q2	The component associated with the rotation about the Y-axis	4
	orientation change q3	The component associated with the rotation about the Z-axis	4
Quaternion	*quaternion q0*	The scalar (real) part of the quaternion, representing the rotation's magnitude or angle	4
	quaternion q1	The component associated with the rotation about the X-axis	4
	quaternion q2	The component associated with the rotation about the Y-axis	4
	quaternion q3	The component associated with the rotation about the Z-axis	4
Euler Angles (XYZ Earth fixed type)	*Roll*	The rotation of an object around its longitudinal axis, which is an axis running from the front to the back of the object	4
	Pitch	The rotation of an object around its lateral axis, which is an axis running from side to side	4
	Yaw	The rotation of an object around its vertical axis, which is an axis running vertically, typically perpendicular to the ground	4
Total count			128

The count represents the multiplication of these four statistics with the features listed in the feature column.

However, unlike gaze analysis, which has established specific features linked to cognitive processes, there is limited literature on how head movements correlate with deeper cognitive tasks. While we can make assumptions—such as variation in Yaw rotation indicating search behavior or changes in walking acceleration reflecting stops or speed adjustments due to engagement in another task (e.g. looking at a map)—we aimed for a more complete encoding of movement. Therefore, we utilized all the precise IMU measurements we had and computed basic statistical features (minimum, maximum, average, and variance) from them. We believe this detailed encoding of the movement will help the model distinguish the movements more effectively. Table 2 summarizes these 128 features derived from IMU recordings.

5. Analysis

This section outlines the analysis and methods applied to address our research questions. As noted in Section 1, the required analysis is structured into two parts: first, training an ML model for predicting the three well-defined wayfinding steps (i.e. the macro steps), and second, decoding the monitoring phase using this trained model. The following subsections explain these parts in detail.

5.1. Part I: Prediction of the macro steps in wayfinding

The first part of our analysis involves training a model to predict the macro steps of wayfinding. Based on our experience with various ML classifiers and supported by the literature, we focus on tree-based classifiers, which are effective at splitting the feature space and are among the best for medium-sized structured data (Treboux *et al.* 2018). Given its promising results in human activity recognition, we selected XGBoost for our experiments (Ambati and El-Gayar 2021). Additionally, we tested an MLP (Multilayer Perceptron) due to its simplicity and effectiveness with structured data, avoiding the complexity of deeper networks that may overfit our dataset size. We conducted experiments with both classifiers. Our goal was to train a robust model that could reliably infer the micro steps during the monitoring phase.

5.1.1. Preparing the data for training and inference

As shown in Figure 4, each instance was labeled based on events logged by the experimenter app. Instances not related to wayfinding activities were excluded, such as those after self-localization (light green bars labeled as 'standing' in Figure 4), when the experimenter interacted with the participant to check the pinned location, ask the orientation question, and display the next destination. These instances are not representative of wayfinding activities and, thus, were omitted from the analysis. The orientation task was also excluded, in line with Downs and Stea's theory, which treats orientation as part of self-localization rather than a separate step.

For goal recognition, participants pressed a clicker upon recognizing the destination and continued walking towards it. If they forgot, the experimenter asked when they had recognized the destination and manually labeled that moment as the clicker event. Since decision-making occurs before clicking, we defined 5, 10, and 15 seconds prior to the clicker press as possible ground truths. Testing revealed that the 10-second interval yielded the highest classification scores, so we adopted this interval as ground truth.

As discussed in Section 1, defining accurate ground truth for the 'monitoring' step is challenging due to its abstract definition. Therefore, we trained the model on the three well-defined macro-steps: self-localization, route planning, and goal recognition (as depicted in Figure 5). This means we excluded any data related to the monitoring part—the segment after route planning, where participants started walking towards the destination until recognizing it—from the model training (marked in red in Figure 5). This segment of data from each route was reserved for inference, allowing us to infer the learned patterns of the macro-steps of wayfinding during the monitoring step.

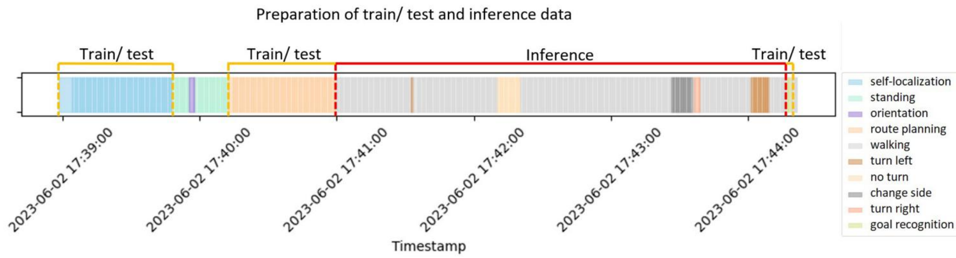


Figure 5. The model is trained and tested on three steps: self-localization, route planning, and goal recognition. The remaining data, representing monitoring (the walk to the destination) is used for inference.

5.1.2. Machine learning experiments

Several ML experiments were conducted using the two models, testing different feature combinations: gaze features only, IMU features only, and both combined. Each experiment was repeated with 18 sliding window approaches. We retained routes from 10 random participants for testing ($\approx 19\%$ of the data), excluding their data from training and validation (i.e. using the Leave-k-Group-Out method). Using 10-fold cross-validation on the remaining data, we tuned the models' hyperparameters. Consistent testing was ensured by using the same 10 left-out participants for all experiments. Given the class imbalance, particularly with goal recognition being represented by only 5 instances per route, we applied the SMOTE (Synthetic Minority Oversampling Technique) upsampling method to balance the dataset for training (Chawla *et al*. 2002). The best model was XGBoost with hyperparameters: *[subsample: 0.5, n-estimators: 4500, max-depth: 7, learning-rate: 0.01, colsample-bytree: 0.45]*, trained on both IMU and gaze features with a 6-second overlapping window segmentation. We use this trained model for inference as described in Section 5.2.

5.2. Part II: Decoding monitoring through sequence analysis

The trained model classifies three of the steps proposed by Downs and Stea. To account for monitoring as an independent step or other undefined steps, we added a fourth class, 'unknown' for predictions below a confidence threshold determined through a data-driven approach. We analyzed class probabilities for the test set, identifying correct predictions and calculating the 10th percentile of their probabilities. This threshold, found to be 67.9%, means 90% of the correct predictions had probabilities above this value, reflecting the model's confidence. Predictions with a probability below this threshold were labeled 'unknown' with the low-confidence predicted class label still noted. Using this trained model, we inferred the monitoring phase which is considered the most ambiguous part of the model by Downs and Stea. Therefore, for each route's monitoring step, we received a sequence of micro steps: self-localization, route planning, goal recognition, and unknown, resulting in 309 sequences (check Section 6.2 for an example of such a sequence). Figure 6 (Section 6) illustrates an example. To derive insights, we applied sequence analysis techniques such as sequential pattern mining and examining label transitions and distributions.

5.2.1. Sequential pattern mining on micro-steps

Sequential pattern mining is a method used to discover recurring sequences or order of items within ordered sets of events or items (Fournier-Viger *et al*. 2017). It is valuable in contexts where the order of occurrences is critical, such as analyzing customer purchase behavior, web navigation paths, or biological sequences. This technique identifies frequent sequences in a dataset, revealing hidden patterns valuable for different analyses. Common algorithms for this task include FPGrowth (Han and Pei 2000), Apriori (Agrawal *et al*. 1994), GSP (Generalized Sequential Pattern) (Srikant and Agrawal 1996), PrefixSpan (Han *et al*. 2001), etc. Among these, we tested three methods: Apriori, FPGrowth, and PrefixSpan. PrefixSpan stands out for its efficiency in handling large datasets without generating and testing an excessive number of candidate sequences. By breaking the problem into smaller parts (prefixes and projected databases), PrefixSpan efficiently identifies frequent patterns, improving mining speed and accuracy (Han *et al*. 2001). Of the three methods tested, we selected PrefixSpan for its ability to generate a more complete set of subsequences, making it the most suitable for our dataset of 309 sequences with varying item counts. The results from PrefixSpan enabled us to identify the most frequent sequence orders across all routes.

5.2.2. Analyzing the share and transition of micro-steps

The second analysis examined the transitions between micro-steps, such as how labels in a sequence like *[a, a, b, c, b, d, a, …]* change over time. This involved analyzing, first, the share of each label in each sequence and, subsequently, how each label transitions to another, e.g. from 'a' to 'b'. We also looked at the distribution of different labels to understand if instances of one label are clustered together or spread out. For example, are instances of label 'a' mostly clustered at the beginning of the sequence, or are they scattered throughout the sequence? The analysis focused on three key aspects: *First*, the share of each micro-step, i.e. the overall time spent on each micro-step. For example, how much time did the participants spend on self-localization during the monitoring? *Second*, the relationship between the time spent on micro self-localization and route planning, and their corresponding macro-steps. *Third*, the distribution of time spent on each step throughout the monitoring period. This means examining how the time for each label is spread out.

6. Results

6.1. Part I: Prediction of the macro steps in wayfinding

Table 3 compares the accuracy and kappa values of MLP and XGBoost models on test data, trained with both IMU and gaze features. XGBoost consistently outperforms MLP across all window sizes. Figure 7 shows that while gaze features alone have limited predictive power, its combination with IMU features achieved the highest accuracy, using a 6-second overlapping sliding window segmentation. To ensure the generalizability of the results, the XGBoost model experiments with both sensor features were repeated five times using five different 10-participants-out method. The results in Table 3 represent the average of these runs. The best-performing model from these runs was used for the sequence analysis.

Figure 8 includes (A) the confusion matrix, (B) the learning curve, and (C) the SHAP (SHapley Additive exPlanations) values (Lundberg and Lee 2017) showing the top 20 important features for predicting each class. The confusion matrix (A) visualizes the model's performance by showing which classes are confused with each other. It shows that goal recognition is almost perfectly predicted, while self-localization and route planning are misclassified as each other for 17 and 18% of the time, highlighting areas where the model struggles. The bars in the SHAP plot (C), color-coded by class, illustrate the importance of each feature for predicting each class. These plots are for the best-performing model tested on unseen test data.

6.2. Part II: Decoding monitoring through sequence analysis

As explained in Subsection 5.2, we used the trained model with a prediction probability threshold for inference. Figure 6 shows an example of the inferred classes within the monitoring step, color-coded by class. The gray color represents the 'unknown' class but still indicates the potential label the model could assign, regardless of prediction confidence. The next two subsections present the results from

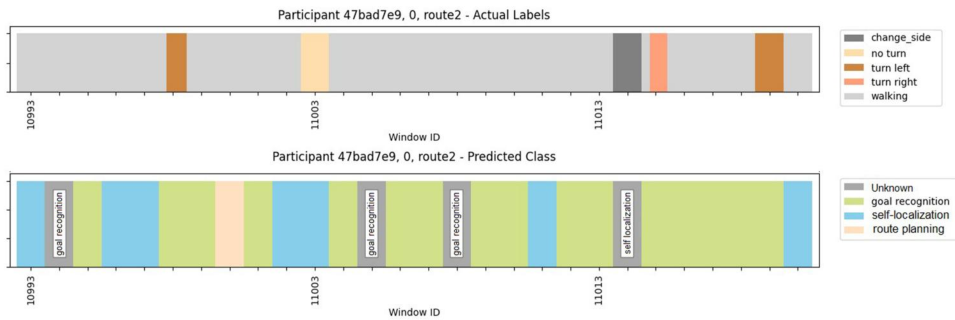


Figure 6. Shows the inferred labels from the trained model. The top plot displays the original events logged by the experimenter during the monitoring part of a sample route, while the bottom plot shows the same part with the inferred labels. The gray color represents the "unknown" class, and the label on it indicates the label the model could have assigned it to if the prediction confidence had not been considered.

Table 3. Test results of the MLP and XGBoost models, trained on both IMU and gaze features. XGBoost outperforms the MLP in all window sizes.

	MLP		XGBoost	
Window-size	Accuracy	Kappa	Accuracy	Kappa
2	0.717	0.544	0.748	0.599
3	0.720	0.540	0.808	0.675
4	0.738	0.574	0.797	0.670
5	0.736	0.581	0.807	0.652
6	0.739	0.572	**0.878**	**0.819**
7	0.705	0.444	0.812	0.712
8	0.719	0.481	0.812	0.640
9	0.729	0.498	0.779	0.598
10	0.641	0.342	0.688	0.519

The reported values for the XGBoost experiments are the average of five runs with five different test sets (i.e. five runs with the Leave-10-Group-Out method).
The highest performance metrics are highlighted in bold.

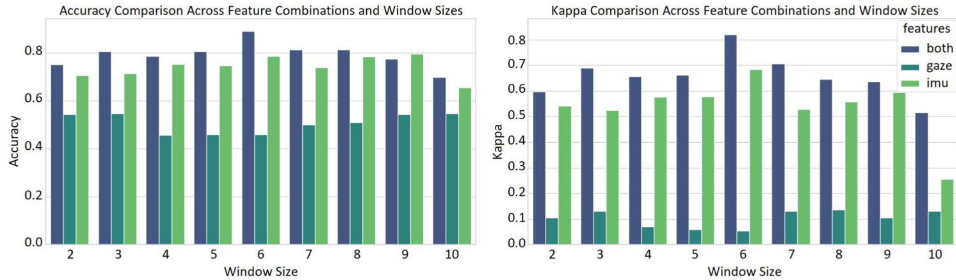


Figure 7. The XGBoost model's performance was compared across different feature combinations and segmentation window sizes. The best results in accuracy and kappa were achieved with both sensors and a 6-second window size.

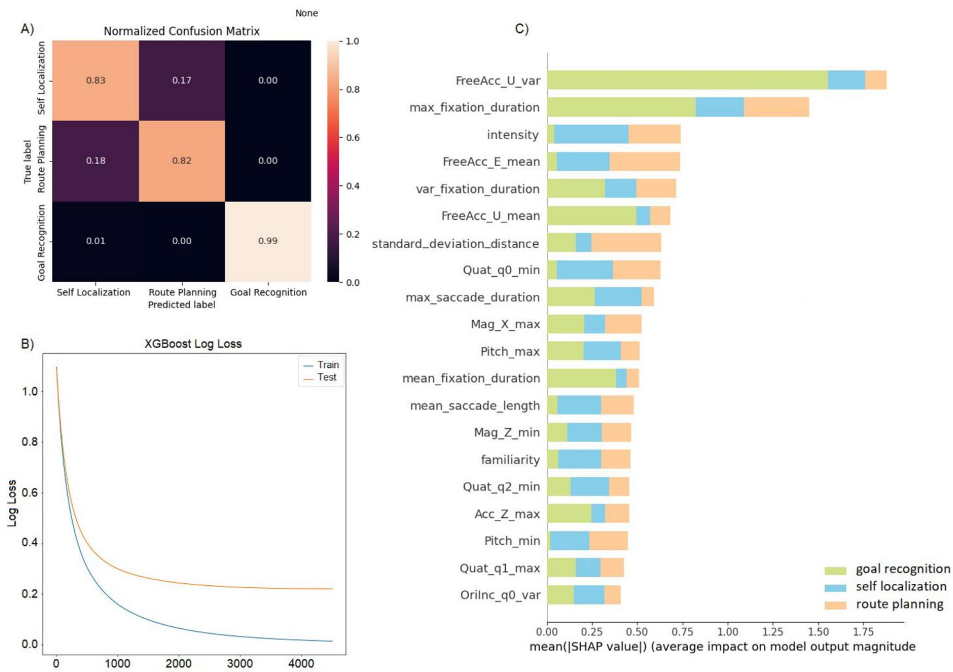


Figure 8. (A) Presents the confusion matrix showing the model's performance for each class. (B) depicts the model's learning curve, illustrating its performance over training time. (C) Displays the SHAP values for the top 20 most important features, color-coded by class, to indicate which features are most effective for predicting each class.

sequential pattern mining and micro-step transition analysis, following the structure in Subsection 5.2.

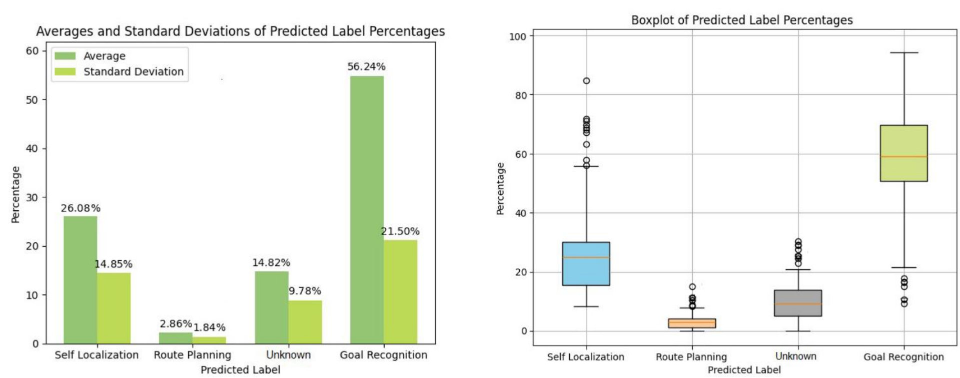
6.2.1. Sequential pattern of micro-steps during monitoring

Sequential pattern mining using the PrefixSpan algorithm identified [*'unknown'*, *'self-localization'*, *'goal recognition'*] as the most frequent subsequence of labels, with a support of 0.915. Input sequences were pre-processed to merge consecutive repetitions of the same label, e.g. [*'a'*, *'a'*, *'a'*, *'b'*, *'c'*, *'c'*] was converted to [*'a'*, *'b'*, *'c'*]. Since our inputs could potentially contain repetitions of these subsequences ([*‘a’*, *‘b’*, *‘c’*, … ,*‘a’*, *‘b’*, *‘c’*, …]), we count the occurrence of each subsequence in each route and calculated support as the total number of occurrences divided by the total number of routes.

6.2.2. The share and transition of micro-steps during monitoring

In Subsection 5.2.2, we posed three questions about micro-step transitions: (1) the overall time spent on each micro-step during monitoring, (2) the relationship between macro and micro self-localization and route planning times during monitoring, and (3) how the time spent on micro-steps is distributed over the monitoring step. We address these questions with appropriate plots. To answer question (1), we use boxplots per label (Figure 9), which show the share of each inferred label during the monitoring across all routes. On average, 56.24% of the monitoring part is labeled as goal recognition, 26.08% as self-localization, 2.86% as route planning, and 14.82% as unknown.

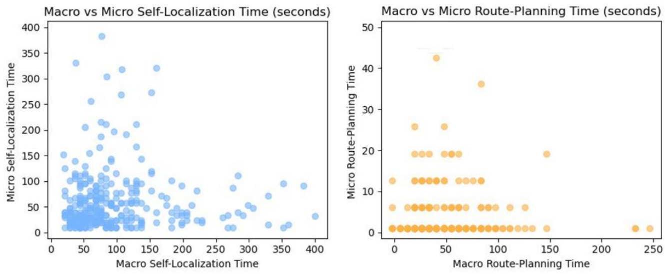
Question (2) is answered by plotting the total micro self-localization and route planning times against their respective macro times. The scatterplots (see Figure 10(a)) reveal distinct patterns for these wayfinding steps. For self-localization (left), most points cluster in the lower left, indicating that both macro and micro times are typically under one minute, though some points show occasional differences. For route planning (right), most points cluster along the x-axis, indicating that micro route planning times are generally short, even with varying macro times (note the different scales on the y and x axes). Question (3) was answered by plotting the time distribution of the micro-steps within the monitoring step (see subplot 10b). The bars represent the total monitoring time, while the circular markers indicate each instance of the micro-step and the unknown class. The y-axis shows the time spent walking to the destination, and the markers reveal the time-wise distribution of micro-steps throughout monitoring. Regardless of the high variability between routes across micro-steps, goal recognition occurs more frequently along the walk; self-localization, although repeated at variable intervals during monitoring, mainly occurs at



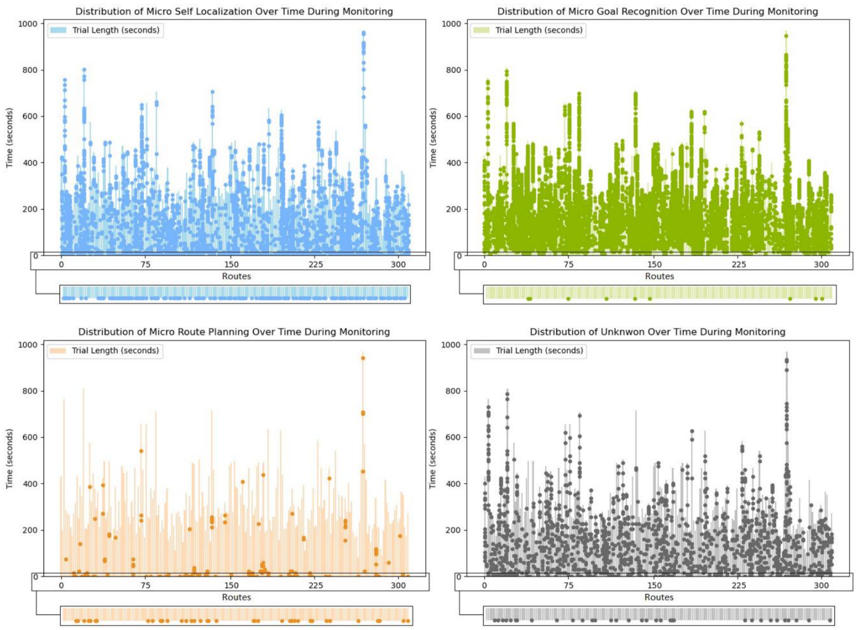
(a) This subplot compares the average and standard deviation of the four inferred labels.

(b) Boxplot of each inferred label showing the median values, inter-quartile ranges, and outliers.

Figure 9. Shows the distribution of micro steps that form the monitoring step. For example, based on the median (b), monitoring consists of 58.76% micro goal recognition, 24.88% micro self-localization, 2.80% route planning, and 12.75% unknown.



(a) The scatterplots show the total time of macro and micro self-localization (left) and route planning (right) for each route. Each marker represents a route, with the total time of macro self-localization plotted against all micro self-localizations, and similarly for route planning. The plots suggest that micro and macro route planning are independent, while micro and macro self-localization exhibit a weak negative correlation.



(b) The plots show the time distribution of each micro-step and the unknown class during monitoring across all routes. Bars indicate the total monitoring duration, and markers represent each inferred class along this timeline. The area around the 0-second mark is zoomed in to highlight initial labels during the start of the monitoring step. These plots show that micro self-localization and goal recognition are the major parts of monitoring.

Figure 10. These plots show the relationship between the macro (initial) and micro (inferred during monitoring) wayfinding steps.

the trip's start; route planning follows a similar pattern on a smaller scale, i.e. mostly at the start of monitoring phase.

7. Discussion

To address our main research question of whether monitoring is a standalone step, as defined by Downs and Stea (1977), or a recursive call of self-localization, route planning, and goal recognition, we first trained a model using behavioral data from the macro version of these three steps (i.e. initial self-localization and route planning, and final goal recognition). Our results show that these steps can be predicted with high accuracy using head/body and eye movement data. The SHAP feature importance analysis revealed that certain behaviors are strongly associated with specific steps. Next, we used this trained model to infer the same behavioral patterns during the monitoring phase across all routes. Our findings suggest that monitoring likely involves micro versions of the same steps. The following subsections provide a detailed discussion of these findings and the remaining open questions.

7.1. Prediction of the three macro steps in wayfinding

The accuracy and kappa metrics reported in Subsection 6.1, indicate that predicting wayfinding steps through ML techniques using visual attention and physical movement features is feasible. XGBoost outperformed MLP, highlighting the strength of ensemble learning in capturing the complexity and multi-dimensional nature of sensory data during wayfinding tasks as also noted by Wu *et al.* (2021). Figure 7 shows that combining gaze and IMU features results in optimal performance, indicating that both visual attention and physical movements are influenced by the underlying cognitive processes of wayfinding (See e.g. Giannopoulos (2016), Takemiya and Ishikawa (2011)).

The SHAP value plot (Figure 8(B)) highlights the importance of these features in predicting the three macro steps. Notably, half of the top 10 features are gaze-related, emphasizing the crucial role of visual attention in wayfinding. Fixation duration, reflecting the cognitive load of information processing (Irwin 2013), as well as saccade length and duration, reflecting task difficulty (Xin *et al.*2021), enhance the model's predictive power almost equally for all three classes. This suggests that each class has a distinct pattern of fixation duration and saccade length and duration, indicating different cognitive processes and visual search strategies (careful orientation vs overview scans) for each step (Alinaghi and Giannopoulos 2024). As expected, the standard distance deviation (stdd) and intensity of fixation distribution are more important for self-localization and route planning, supporting the idea that gaze distribution differs during these steps (Kiefer *et al.*2013). During route planning, fixations tend to be linearly clustered along the planned route, whereas during self-localization, they are more dispersed across the tablet and surrounding environment. These variations in gaze patterns can also reflect changes in cognitive load. Clustered gaze during route planning shows higher cognitive demand, while dispersed gaze during self-localization reflects less cognitive demand (Nakayama *et al.*2002).

While five gaze features are among the top 10, the model's performance significantly dropped when trained only on gaze data. Although eye movements are task-related and, therefore, provide good proxies for understanding the cognitive processes of a task, they are also influenced by external factors, such as noise. Also, research has shown that, for example, during periods of mind wandering, gazing somewhere does not necessarily indicate an active perception of information (See e.g. Kwok *et al.*

(2024)). Combining eye tracking and behavioral sensors, hence, improves prediction accuracy as indicated by our empirical evidence.

The computed IMU features capture various aspects of the head/body movements. Among these features, the variance of free acceleration along the Z-axis (FreeAcc-U-var) is the most important feature particularly for predicting goal recognition, as it encodes variations in walking and gait changes. The IMU's Z-axis faces upward, capturing upward movements associated with the swing phase of gait while walking (Gujarathi and Bhole 2019). The confusion matrix confirms the feature's strong predictive power for goal recognition. The average free acceleration along the X-axis (FreeAcc-E-mean), pointing East, is key for detecting route planning and self-localization, as it reflects lateral velocity changes in head/body movements during activities like looking around. Pitch-related features (i.e. up-and-down head movements) are more effective for predicting self-localization and route planning but less for goal recognition, indicating minimal 'looking up-and-down' during goal recognition.

7.2. Monitoring: a standalone step or a recurring call of micro-steps?

Using the trained model and applying sequence analysis, we decoded the monitoring step, uncovering a sequence of predicted labels at a micro level. This shows that the same physical and gaze behavior patterns observed during self-localization, route planning, and goal recognition are also present during monitoring. Cognitive neuroscience research has linked cognitive processing to brain-behavior relationships, showing that, for example, during perceptual decision-making, the decision-making components at two representational levels (neural and behavioral) are significantly associated (Imani *et al*. 2021). Therefore, in our study, the recurrence of similar behavioral patterns during monitoring can also suggest a connection to similar underlying cognitive processes. Cognitive processing, broadly defined, involves a sequence of stages in which sensory input is transformed, reduced, elaborated, stored, retrieved, and utilized (Kreutzer *et al*.2011) [p. 859]. Considering this, monitoring serves as a concrete example of cognitive processing in action. By continuously iterating through micro self-localization, micro route planning, and micro goal recognition, monitoring illustrates how cognitive operations such as attention, memory, and reasoning dynamically interact to support wayfinding. This recursive process highlights the adaptive and non-linear nature of cognitive processing. The following subsections explain the role of each micro-step in 'keeping us on the right track'.

7.2.1. The role of micro self-localization

As shown in Figure 9, self-localization is the second most frequent label, accounting for 26.08% of the monitoring time. But how does micro self-localization differ from, or resemble, macro self-localization? Micro self-localization is consistently distributed throughout the monitoring phase, including near the destination (Figure 10(b)), supporting findings that individuals reaffirm their position as they move farther from the starting point (Wiener *et al.* 2009). Early in monitoring (in the zoomed-in view in Figure 10(b)), 58.22% of participants spent the first six seconds on self-localization, indicating active recapping of their position while walking, consistent with Montello and Sas (2006)'s concept of orientation updating. Figure 10(a) shows a weak negative correlation (-0.202, $p = 0.0277$) between the time spent on macro and micro self-localization, as revealed by a Spearman's rank correlation test. This statistically significant but weak relationship suggests that more time spent on macro self-localization reduces the need for micro self-localization fine-tuning, aligning with Passini (1984), who noted that a detailed mental representation reduces the need for frequent recall during navigation. This, however, needs further analysis regarding the working memory of the subjects. Another aspect could be the impact of spatial scale on the cognitive functions triggered for any step (Wolbers and Wiener 2014). Macro self-localization occurs in vista space, i.e. where all necessary information is visible within a field of view with minimal movement, whereas monitoring happens in environmental space, i.e. where movement is required to access necessary information. However, according to Montello (1993), environmental space can be viewed as a connected stream of vista spaces, therefore suggesting micro self-localization takes place across multiple such discrete spaces. Together with a cognitive view, micro self-localization combines real-time sensory input (bottom-up processing) with spatial knowledge (top-down processing) to determine a person's position in their environment. This process reflects the mechanisms of macro self-localization, but works continuously and adaptively in the dynamic and varying landscape of environmental space.

7.2.2. The role of micro route planning

Route planning is the least frequent label, accounting for only 2.86% of the monitoring time (Figure 9). As shown in Figure 10(a), the micro route planning time is roughly one-fifth of the macro time, with several cases showing no instance of micro route planning at all. This could be due to the relatively short routes, allowing participants to easily memorize them after initial planning. This interpretation aligns with Alinaghi *et al*. (2023), who discuss that longer distances increase the likelihood of forgetting instructions and needing them again. It is also consistent with the serial position effect, where longer sequences make intermediate items harder to recall (Ebbinghaus 1913). Hilton *et al*. (2021) have found a trace of this effect in route learning by memorizing landmarks. They have also reported that participants break longer sequences into smaller sub-lists. This observation could explain the behavior of revisiting route planning during monitoring. Routes with higher complexity (e.g. longer distances, more decision points) are likely to impose greater working memory demands, leading to more frequent instances of micro self-localization and route planning. Another explanation could be the distinction Hölscher *et al*. (2011) makes between planning a route for oneself before taking it (prospective planning) and planning a route while navigating (situated planning). According to their study on familiar wayfinders, situated planning can be seen as 'an incremental optimization of the overall plan by adding in local direction information'. Our inferred micro route planning could be an instance of situated planning. Despite its minimal overall occurrence, 20.39% of micro route planning occur when participants start walking, possibly as a rehearsal of the planned route.

7.2.3. The role of micro goal recognition

According to Figure 9, on average, 56.24% of the monitoring phase is dedicated to goal recognition, with a high standard deviation, indicating variability across routes that could stem from individual differences. The boxplots show that despite this variability, goal recognition constitutes the majority of the monitoring phase for all routes. Figure 10(b) reveals that goal recognition occurs consistently throughout the walk to the destination. This finding aligns with problem-solving theories, which suggest individuals set intermediate sub-goals to monitor progress, adjusting their actions continuously until the final goal is achieved (Kaplan and Kaplan 1982, MacGregor *et al.* 2001). Goal recognition's predominance during monitoring can also align with working memory's role in comparing stored goal representations with sensory input to assess progress (Ebbinghaus 1913). Despite the fact that goal recognition is the most frequent step throughout the monitoring phase, it accounts only for 2.9% of the first six seconds. This suggests that even though macro self-localization and route planning were performed immediately prior (i.e. the person found their current location and planned a route to the destination just a few seconds before), goal recognition is not the first cognitive function to be activated during the monitoring phase. Considering the three levels of situational awareness—perceiving, comprehending, and projecting future states in the environment (Endsley *et al*. 2000)—goal recognition corresponds to the third level, such as anticipating landmarks or assessing progress toward the destination. This suggests that participants were likely still focused on the first two levels of situational awareness at the beginning of the monitoring phase.

7.2.4. What can 'unknown' mean?

Figure 9 shows that 14.82% of the monitoring time could not be confidently classified by the model. This is a significant observation, as this unclassified time is almost evenly distributed throughout the monitoring. In particular, in 18.48% of all routes, the monitoring begins with six seconds of this 'unknown' step (Figure 10(b)). Several potential explanations come to mind: First, it could represent a distinct and standalone step as suggested by Downs and Stea (1977), possibly a pure control mechanism to ensure the wayfinder stays on track. Second, it might indicate other activities or mental states, such as mind wandering (Lee *et al.* 2021). Third, it could reflect transitional periods between cognitive steps. While cognitive processes are often modeled as discrete steps for analysis, they are typically continuous and dynamic, with overlapping stages and gradual transitions. This also aligns with Downs and Stea (1977)'s conceptualization of the steps as interrelated. In psychology, it is well-known that cognitive tasks often transition into one another, with one process gradually giving way to the next (Spivey 2007, p. 3–29). Further supporting the idea of transitions, we observed that the 'unknown' label never appeared consecutively in the inferred sequences—other labels always occurred between instances of 'unknown'. This supports the third explanation more as a single unknown prediction is too short to be a pure control mechanism, and the consistent distribution makes it less likely to be another activity, e.g. mind wandering or window shopping. Our assumption is that these unknown instances are more likely to represent moments of transition between steps, where the model struggles to classify them into one of the predefined classes due to the mixture of steps involved. With this interpretation, 'unknown' likely represents moments of heightened cognitive load as individuals integrate information across steps. These transitions align with theories of dual-task interference, where concurrent cognitive processes (e.g. updating spatial information and planning) compete for limited cognitive resources (Spivey 2007, p. 98).

7.2.5. Ordered sequence of wayfinding steps

So far, we have primarily focused on each step individually. However, an essential aspect of the theory by Downs and Stea is the sequence of these steps. Do the steps follow the order defined in the theory, or do they occur without a specific sequence? While our experimental design dictates a certain order for the first two measured steps, the PrefixSpan algorithm can help explore whether such an order can be extracted from our empirical analysis of the monitoring step. The resulting sequence for the monitoring phase –[*'unknown'*, *'self-localization'*, *'goal recognition'*]– suggests that, regardless of the number of occurrences of these three labels, this order is observed in 91.5% of all cases. It is important to note that this order does not imply any hierarchy between the steps, but rather reflects the most frequently observed order across routes. This suggests that during the walk to the known destination, participants encounter an initial unknown step, potentially reflecting a transitional moment after completing the macro route planning. This is followed by micro-level self-localization and then micro-level goal recognition. This cycle with this order can repeat throughout the monitoring phase.

8. Conclusion

In this study, the cognitive processes involved in wayfinding were empirically investigated with a focus on the monitoring step, based on the model proposed by Downs and Stea (1977). We trained a ML model using behavioral data from 56 participants navigating over 300 routes. The model successfully learned behavioral patterns related to three of the wayfinding steps: self-localization, route planning, and goal recognition, achieving a test accuracy of 87.8%. The trained model was then used to infer the learned behavioral patterns during the monitoring phase of all routes. Sequence analysis on the inferred classes revealed that monitoring likely involves recurring iterations of these three steps, suggesting that monitoring is a phase of continuous mid-goal setting and fine-tuning of self-localization and route planning.

Despite these insights, some limitations and questions for future research remain. The gender imbalance in the sample could not be addressed during recruitment but may have an impact on the findings. Furthermore, this study focused on predicting the four steps and analyzing the monitoring phase based solely on gaze behavior and head/body movements. However, existing literature indicates that personal attributes, spatial familiarity, and environmental features significantly impact wayfinding behavior (Giannopoulos *et al.* 2014). Future research should explore how these factors influence the interpretation of monitoring. For instance, how do familiarity and spatial skills affect the results? How do familiar and unfamiliar wayfinders perform in different steps? Do familiar participants have more efficient self-localization or route planning strategies, or is their performance more related to spatial skills (Kattenbeck *et al.* 2024)? As an example of environmental factors, do street junctions, commonly considered as decision points, correlate with goal recognition behavior? The unknown steps require more in-depth analysis, particularly regarding the semantic information in visual attention. Previous research by Alinaghi and Giannopoulos (2024) has identified distinct patterns in visual attention across the four steps of wayfinding. It is crucial to explore the semantic information related to visual attention—specifically, the objects that wayfinders focus on—during these instances of unknown prediction. Furthermore, while we focused on the four steps suggested by Downs and Stea, it is worth exploring whether more and/or different steps exist. Employing unsupervised learning techniques, which do not require ground truth data, could uncover additional patterns in the collected behavioral data.

Ultimately, this study enhances our understanding of human wayfinding behavior, informing the design of more efficient and accessible navigational tools and urban environments. Such empirical studies are crucial for building new theories or refining existing ones, advancing our theoretical and practical understanding of spatial cognition and wayfinding. For instance, our findings suggest that monitoring might involve iterative cycles between the steps of self-localization, route planning, and goal recognition, hinting at the dynamic and interrelated nature of the wayfinding steps proposed by Downs and Stea (1977).

Notes

1. The experimental design proposal was reviewed by the Pilot Research Ethics Committee of Vienna University of Technology(see supplementary materials). All participants have given written consent prior to participation in the study.
2. While not all the collected data was used in this analysis, we present a comprehensive overview of the data collection procedure to help readers understand why and how specific data were gathered.
3. The dropout rate may have been due to the longer and physically demanding outdoor sessions compared to the online step.
4. The directions *X*, *Y*, and *Z* correspond to *East*, *North*, and *Up*, respectively.

Acknowledgments

We sincerely thank all participants for their time and effort in supporting our research. We also acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Programme. We acknowledge the use of AI-assisted tools (Premium Writefull extension of Overleaf) for final proofreading and refining the language of this article.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Negar Alinaghi is a PhD candidate in the Geoinformation Group at TU Wien. Her research focuses on Human Activity Recognition and GeoAI.

Ioannis Giannopoulos is a Full Professor of Geoinformation at TU Wien. His research interests focus on Urban Computing, GeoAI, Mobile and Remote Eye Tracking, Human Mobility, Pedestrian Navigation, and Mixed Environments.

Markus Kattenbeck is a Senior Scientist within the Geoinformation Group at TU Wien. He has been the head researcher of the Spatial-HCI Lab since 2019. His research interests are highly interdisciplinary and range from spatial cognition, human wayfinding, and its behavioral correlates to Human-Computer Interaction within and beyond Augmented Reality.

Martin Raubal is a Professor of Geoinformation engineering at ETH Zurich. His research interests focus on spatial decision-making for sustainability; more specifically, he concentrates on analyzing spatiotemporal aspects of human mobility, spatial cognitive engineering, and mobile eye-tracking to investigate visual attention while interacting with Geoinformation and in spatial decision situations.

ORCID

Negar Alinaghi http://orcid.org/0000-0001-6594-4481
Ioannis Giannopoulos http://orcid.org/0000-0002-2556-5230
Markus Kattenbeck http://orcid.org/0000-0001-6028-0428
Martin Raubal http://orcid.org/0000-0001-5951-6835

Data and codes availability statement

The data and codes that support the findings of this study are available at the public link: https://doi.org/10.48436/m2ha4-t1v92.

References

Agrawal, R., and Srikant, R., *et al.*, 1994. Fast algorithms for mining association rules. *In*: *Proceedings of 20th International Conference of very large data bases*. Santiago: VLDB, vol. 1215, 487–499.

Alinaghi, N., and Giannopoulos, I., 2022. Consider the head movements! Saccade computation in mobile eye-tracking. *In*: *2022 Symposium on Eye Tracking Research and Applications*, Seattle, WA, USA. New York, NY: Association for Computing Machinery, 1–7.

Alinaghi, N., and Giannopoulos, I., 2024. Wayfinding stages: The role of familiarity, gaze events, and visual attention. *In*: *16th International Conference on Spatial Information Theory (COSIT 2024)*. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 1–1.

Alinaghi, N., Kattenbeck, M., and Giannopoulos, I., 2022. I can tell by your eyes! Continuous gaze-based turn-activity prediction reveals spatial familiarity. *In*: *15th International Conference on Spatial Information Theory (COSIT 2022)*. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

Alinaghi, N., *et al.*, 2023. Do you need instructions again? Predicting wayfinding instruction demand. *In*: *GIScience 2023*. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

Ambati, L.S., and El-Gayar, O., 2021. Human activity recognition: a comparison of machine learning approaches. *Journal of the Midwest Association for Information Systems*, 2021 (1), 4.

Arthur, P., and Passini, R., 1992. *Wayfinding: people, signs, and architecture.* New York: McGraw-Hill Book Co.

Baddeley, A., Rubak, E., and Turner, R., 2016. *Spatial point patterns: methodology and applications with R*. Volume 1. Boca Raton, FL: CRC Press.

Brügger, A., Richter, K.F., and Fabrikant, S.I., 2018. Which egocentric direction suffers from visual attention during aided wayfinding? *In*: *Eye Tracking for Spatial Research, Proceedings of the 3rd International Workshop*. Zurich, Switzerland: ETH Zurich.

Carver, C.S., and Scheier, M.F., 2001. *On the self-regulation of behavior*. Cambridge: Cambridge University Press.

Chawla, N.V., *et al.*, 2002. Smote: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357.

Chown, E., Kaplan, S., and Kortenkamp, D., 1995. Prototypes, location, and associative networks (plan): towards a unified theory of cognitive mapping. *Cognitive Science*, 19 (1), 1–51.

Dalton, R.C., 2003. The secret is to follow your nose: route path selection and angularity. *Environment and Behavior*, 35 (1), 107–131.

Dehghani, A., *et al.*, 2019. A quantitative comparison of overlapping and non-overlapping sliding windows for human activity recognition using inertial sensors. *Sensors*, 19 (22), 5026.

Dong, W., *et al.*, 2022. Wayfinding behavior and spatial knowledge acquisition: Are they the same in virtual reality and in real-world environments? *Annals of the American Association of Geographers*, 112 (1), 226–246.

Downs, M.R., and Stea, D., 1977. *Maps in minds: reflections on cognitive mapping. Harper & Row Series in Geography*. New York, NY: Harper & Row.

Duchowski, A.T., 2017. *Eye tracking methodology: Theory and practice*. Cham, Switzerland: Springer International Publishing AG.

Ebbinghaus, H., 1913. *Memory: a contribution to experimental psychology. No. 3*. New York: Teachers College, Columbia University.

Endsley, M.R., and Garland, D.J., *et al.*, 2000. Theoretical underpinnings of situation awareness: A critical review. *Situation Awareness Analysis and Measurement*, 1 (1), 3–21.

Epstein, R.A., and Vass, L.K., 2014. Neural systems for landmark-based wayfinding in humans. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369 (1635), 20120533.

Fournier-Viger, P., *et al.*, 2017. A survey of sequential pattern mining. *Data Science and Pattern Recognition*, 1 (1), 54–77.

Giannopoulos, I., 2016. *Supporting wayfinding through mobile gaze-based interaction*. Thesis (PhD). ETH Zurich.

Giannopoulos, I., *et al.*, 2014. Wayfinding decision situations: a conceptual model and evaluation. *In*: *Geographic Information Science: 8th International Conference, GIScience 2014*, September 24–26, 2014*, Proceedings 8*. Vienna, Austria: Springer, 221–234.

Golledge, R.G., 1999. *Wayfinding behavior: cognitive mapping and other spatial processes*. Baltimore, MD: JHU Press.

Golledge, R.G., 2003. Human wayfinding and cognitive maps. *Colonization of unfamiliar landscapes: the archaeology of adaptation*. London, UK: Routledge, 25.

Gujarathi, T., and Bhole, K., 2019. Gait analysis using IMU sensor. *In*: *2019 10th International Conf. on Computing, Communication and Networking Technologies (ICCCNT)*. New York City, NY: IEEE, 1–5.

Han, J., and Pei, J., 2000. Mining frequent patterns by pattern-growth: methodology and implications. *ACM SIGKDD Explorations Newsletter*, 2 (2), 14–20.

Han, J., *et al.*, 2001. Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth. *In*: *Proceedings of the 17th International Conference on data engineering*. New York City, NY: IEEE, 215–224.

Hartley, T., *et al.*, 2003. The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. *Neuron*, 37 (5), 877–888.

Hegarty, M., 2002. Development of a self-report measure of environmental spatial ability. *Intelligence*, 30 (5), 425–447.

Hilton, C., Wiener, J., and Johnson, A., 2021. Serial memory for landmarks encountered during route navigation. *Quarterly Journal of Experimental Psychology*, 74 (12), 2137–2153.

Hölscher, C., Tenbrink, T., and Wiener, J.M., 2011. Would you follow your own route description? cognitive strategies in urban route planning. *Cognition*, 121 (2), 228–247.

Imani, E., *et al.*, 2021. Brain-behavior relationships in the perceptual decision-making process through cognitive processing stages. *Neuropsychologia*, 155, 107821.

Irwin, D.E., 2013. Fixation location and fixation duration as indices of cognitive processing. *In*: *The interface of language, vision, and action*. Oxfordshire, UK: Taylor & Francis Group, 105–133.

Jamshidi, S., and Pati, D., 2021. A narrative review of theories of wayfinding within the interior environment. *HERD*, 14 (1), 290–303.

Just, M.A., and C, P.A., 1976. Eye fixations and cognitive processes. *Cognitive Psychology*, 8 (4), 441–480.

Kang, W., *et al.*, 2023. Pysal/pointpats: pointpats 2.3.0. https://doi.org/10.5281/zenodo.7706219

Kaplan, S., and Kaplan, R., 1982. *Cognition and environment: Functioning in an uncertain world*. New York: Praeger.

Kattenbeck, M., *et al.*, 2024. Is familiarity reflected in the spatial knowledge revealed by sketch maps? *In*: B. Adams, A.L. Griffin, S. Scheider and G. McKenzie, eds. *16th International Conference on Spatial Information Theory (COSIT 2024)*, Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Leibniz International Proceedings in Informatics (LIPIcs), vol. 315, 6:1–6:18.

Kiefer, P., Giannopoulos, I., and Raubal, M., 2013. Using eye movements to recognize activities on cartographic maps. *In*: *Proceedings of the 21st ACM SIGSPATIAL International Conf. on Advances in Geographic Information Systems*, Orlando, Florida. New York, NY: Association for Computing Machinery, 488–491.

Kiefer, P., Giannopoulos, I., and 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.

Kiefer, P., *et al.*, 2017. Eye tracking for spatial research: cognition, computation, challenges. *Spatial Cognition & Computation*, 17 (1–2), 1–19.

Kreutzer, J.S., Caplan, B., and DeLuca, J., 2011. *Encyclopedia of clinical neuropsychology*. vol. 28. New York: Springer.

Kwok, T.C., *et al.*, 2024. Gaze-based detection of mind wandering during audio-guided panorama viewing. *Scientific Reports*, 14 (1), 27955.

Lang, F.R., *et al.*, 2011. Short assessment of the big five: ROBUST across survey methods except telephone interviewing. *Behavior Research Methods*, 43 (2), 548–567.

Lee, H.H., *et al.*, 2021. When eyes wander around: Mind-wandering as revealed by eye movement analysis with hidden Markov models. *Sensors*, 21 (22), 7569.

Lundberg, S., and Lee, S., 2017. A unified approach to interpreting model predictions. *In*: *Proceedings of NIPS 2017*. Red Hook, NY: Curran Associates, Inc., 4768–4777.

Lynch, K., 1964. *The image of the city*. Cambridge, MA: MIT press.

MacGregor, J.N., Ormerod, T.C., and Chronicle, E.P., 2001. Information processing and insight: a process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27 (1), 176–201.

Meilinger, T., *et al.*, 2007. How much information do you need? schematic maps in wayfinding and self localisation. *In*: *Spatial Cognition V Reasoning, Action, Interaction: International Conf. Spatial Cognition 2006*, Bremen, Germany, September 24–28, 2006, *Revised Selected Papers 5*. Bremen, Germany: Springer, 381–400.

Miller, G.A., Eugene, G., and Pribram, K.H., 2017. Plans and the structure of behaviour. *In*: W. Buckley, ed. *Systems research for behavioral science*. London, UK: Routledge, Taylor & Francis, 369–382. https://doi.org/10.4324/9781315130569

Montello, D., and Raubal, M., 2012. Functions and applications of spatial cognition. *In*: D. Waller and L. Nadel, eds. *Handbook of spatial cognition*. Washington, DC: American Psychological Association, 249–264.

Montello, D.R., 1993. Scale and multiple psychologies of space. *In*: A.U. Frank and I. Campari, eds. *European conference on spatial information theory*. Berlin, Heidelberg: Springer, 312–321.

Montello, D.R., 2001. Spatial cognition. *In*: N.J. Smelser and P.B. Baltes, eds. *International encyclopedia of the social & behavioral sciences*. Oxford: Elsevier, 14771–14775.

Montello, D., and Sas, C., 2006. Human factors of wayfinding in navigation. *In*: *International encyclopedia of ergonomics and human factors*. CRC Press/Taylor \& Francis, Ltd. https://doi.org/10.1201/9780849375477.ch394

Nakayama, M., Takahashi, K., and Shimizu, Y., 2002. The act of task difficulty and eye-movement frequency for the 'oculo-motor indices'. *In*: *Proceedings of the 2002 symposium on eye tracking research & applications*, New Orleans, Louisiana. New York, NY: Association for Computing Machinery, 37–42. https://doi.org/10.1145/507072.507080

Passini, R., 1981. Way finding: A conceptual framework. *Urban Ecology*, 5 (1), 17–31.

Passini, R., 1984. Spatial representations, a wayfinding perspective. *Journal of Environmental Psychology*, 4 (2), 153–164.

Peer, M., Nadar, C., and Epstein, R.A., 2023. The format of the cognitive map depends on the structure of the environment. *Journal of Experimental Psychology: General*, 153 (1), 224–240.

Raubal, M., 2001. Human wayfinding in unfamiliar buildings: a simulation with a cognizing agent. *Cognitive Processing*, 2 (3), 363–388.

Reimer, F., *et al.*, 2022. Data description of "building age map", Vienna, around 1920. *Data in Brief*, 41, 107864.

Salvucci, D., and Goldberg, J., 2000. Identifying fixations and saccades in eye-tracking protocols. *In*: P*roceedings of the 2000 Symposium on eye tracking research & applications*, Palm Beach Gardens, Florida, USA. New York, NY: Association for Computing Machinery, 71–78.

Schmid, F., Richter, K.F., and Peters, D., 2010. Route aware maps: multigranular wayfinding assistance. *Spatial Cognition & Computation*, 10 (2–3), 184–206.

Seidel, A., 1982. Way-finding in public spaces: The Dallas/Fort worth, USA airport. *In*: *20th International Congress of Applied Psychology*, Edinburgh, UK. The International Association of Applied Psychology (IAAP), 15–31.

Spiers, H.J., and Maguire, E.A., 2008. The dynamic nature of cognition during wayfinding. *Journal of Environmental Psychology*, 28 (3), 232–249.

Spivey, M.J., 2007. *The continuity of mind*. Oxford: Oxford University Press.

Srikant, R., and Agrawal, R., 1996. Mining sequential patterns: Generalizations and performance improvements. *In*: *International Conference on extending database technology*. Berlin, Heidelberg: Springer, 1–17.

Takemiya, M., and Ishikawa, T., 2011. I can tell by the way you use your walk: Real-time classification of wayfinding performance. *In*: *Spatial Information Theory: 10th International Conf., COSIT 2011*, Belfast, ME, USA, September 12–16, 2011, *Proceedings 10*. Belfast, ME, USA: Springer, 90–109.

Tonsen, M., Baumann, C.K., and Dierkes, K., 2020. A high-level description and performance evaluation of pupil invisible. arXiv preprint arXiv:2009.00508. https://arxiv.org/abs/2009.00508

Treboux, J., Genoud, D., and Ingold, R., 2018. Decision tree ensemble vs. NN deep learning: efficiency comparison for a small image dataset. *In*: *Proceedings of the IWBIS 2018*. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE), 25–30.

Vanrell, S.R., *et al.*, 2018. Assessment of homomorphic analysis for human activity recognition from acceleration signals. *IEEE Journal of Biomedical and Health Informatics*, 22 (4), 1001–1010.

Wiener, J.M., Büchner, S.J., and Hölscher, C., 2009. Taxonomy of human wayfinding tasks: a knowledge-based approach. *Spatial Cognition & Computation*, 9 (2), 152–165.

Wolbers, T., and Wiener, J.M., 2014. Challenges for identifying the neural mechanisms that support spatial navigation: the impact of spatial scale. *Frontiers in Human Neuroscience*, 8, 571.

Wu, J., Li, Y., and Ma, Y., 2021. Comparison of xgboost and the neural network model on the class-balanced datasets. *In*: *2021 IEEE 3rd international Conference on frontiers technology of information and computer (ICFTIC)*. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE), 457–461.

Xin, L., *et al.*, 2021. Detecting task difficulty of learners in colonoscopy: evidence from eye-tracking. *Journal of Eye Movement Research*, 14 (2), 1–13. https://doi.org/10.16910/jemr.14.2.5

Xsens Technologies 2020. *MTI user manual, MTI 10-series and MTI 100-series 5th generation*. Enschede, Netherlands: Xsens Technologies.

Yang, F., Fang, Z., and Guan, F., 2020. What do we actually need during self-localization in an augmented environment? *In*: S. Di Martino, Z. Fang, and K.-J. Li, eds. *International Symposium on web and wireless geographical information systems*. Cham: Springer International Publishing, 24–32.

Yarbus, A.L., 1967. *Eye movements and vision*. New York: Plenum Press.