Int. J. Human–Computer Studies 179 (2023) 103098

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Turn Left Turn Right - Delving type and modality of instructions in navigation assistant systems for people with visual impairments

Bineeth Kuriakose [a,*], Ida Marie Ness [a], Maja Åskov Tengstedt [a], Jannicke Merete Svendsen [a], Terese Bjørseth [a], Bijay Lal Pradhan [b], Raju Shrestha [a]

[a] *Department of Computer Science, Oslo Metropolitan University, Oslo, Norway*
[b] *Department of Statistics, Tribhuvan University, Kathmandu, Nepal*

ARTICLE INFO

Keywords:
Navigation assistant
Blind
Visual impairment
Instruction type
Modality
Audio
Vibrations

ABSTRACT

Receiving navigation directions and relevant information through appropriate channels is crucial for individuals with visual impairments when they use navigation assistant systems. Various navigation assistant systems have implemented diverse methods and modalities to provide navigation instructions to users. Moreover, several studies have examined the preferences of individuals with visual impairments regarding modalities. These studies have primarily relied on surveys or interviews to report their findings. In this study, we extensively investigated the type of navigation information individuals with visual impairments would prefer to receive. To achieve this, we employed a mixed-method design comprising an experiment, a questionnaire, and interviews conducted with participants with visual impairments and individuals with 20/20 vision who were blindfolded. We tested unimodal instruction modes (audio only) and multimodal instruction modes (audio and vibration), along with two types of audio instructions: short and descriptive. To evaluate the participants' navigation performance, we measured the time taken and errors made during navigation using different instruction types and modes. The findings of this study suggest that the instruction mode does not have any significant effect on the time taken or errors made to complete the trials. Additionally, no evidence was found to support the preference for descriptive instructions over short instructions among the participants. Furthermore, the study did not find evidence to suggest that participants prefer receiving detailed instructions about the environment and obstacles during navigation. It became evident that participants had varying preferences for instruction modes, highlighting the importance of customized functionalities in navigation assistant systems. Moreover, our study validated that participants with visual impairments spent less time and made fewer errors during navigation compared to blindfolded participants with 20/20 vision. We believe that the results of this study will contribute to expanding the understanding of user preferences regarding the type and modality of instructions in the domain of assistive navigation technology.

1. Introduction

Myriad navigation systems have been proposed in the literature to assist people with visual impairments (Khan et al., 2021; Kuriakose et al., 2022). Effective and accurate information and instruction are essential to guide the users to reach the destination safely and successfully (Xu et al., 2020; Liu et al., 2021). For sighted persons, vision serves as their primary sensory input while navigating. However, other senses, such as hearing, touch, and smell, also support navigation by providing non-visual information. Different navigation systems have used this knowledge to provide output instructions to people with vision impairments. Navigation systems using only one mode of instruction can be referred to as unimodal, while navigation systems that utilize more than one mode of instruction can be referred to as multimodal (Jaimes and Sebe, 2007). Instruction modes through audio (Xiao et al., 2015; Lee and Medioni, 2016), vibrotactile (Wang et al., 2017; Bouzit et al., 2004), and kinesthetic (Amemiya and Sugiyama, 2010) are the three common modalities that have been utilized in various assistive navigation systems.

Audio instruction modality is comprehended by audio descriptions (Anagnostakis et al., 2016; Firmino and Teófilo, 2013), spatial audio (Holland et al., 2002; Katz and Picinali, 2011), and spatial patterns (Shoval et al., 1994). Audio descriptions can provide turn-by-turn

* Corresponding author.
E-mail addresses: bineethk@oslomet.no (B. Kuriakose), s344233@oslomet.no (I.M. Ness), majateng@oslomet.no (M.Å. Tengstedt), s346207@oslomet.no (J.M. Svendsen), s186649@oslomet.no (T. Bjørseth), bijaya.pradhan@ac.tu.edu.np (B.L. Pradhan), raju.shrestha@oslomet.no (R. Shrestha).

https://doi.org/10.1016/j.ijhcs.2023.103098
Received 3 January 2023; Received in revised form 25 April 2023; Accepted 19 June 2023
Available online 22 June 2023

navigation instructions and enrich the journey with verbal information about the navigation environment. It is important to consider the level of detail carefully when designing a system (Manduchi and Coughlan, 2012). Audio description is a commonly used instruction modality in navigation systems (Kuriakose et al., 2022). Some systems use non-verbal audio, such as spatial and acoustic patterns, to give navigation directions to the users (Wilson et al., 2007). Spatial audio can map the sound source to the target directions and is more intuitive when indicating directions (Spagnol et al., 2018). However, research also indicates spatial audio is unbefitting for high-frequency and continuous instructions during navigation since it might interfere with users' perceptions of acoustic environmental cues (Katz and Picinali, 2011).

Vibrotactile instruction modality is another common form used in navigation systems. Vibration patterns can indicate directions for the user to navigate. Vibrations can be delivered using vibratory patterns by placing the vibrator on body parts such as the head, shoulders, wrists, feet, hands, etc. For example, PocketNavigator (Pielot et al., 2010) can point out navigation cues by encoding their direction and distance in vibration patterns. Some vibration-based navigation systems use a different approach, in which they are designed with specific vibration locations and have different target directions. ActiveBelt (Tsukada and Yasumura, 2004), and Tactile Wayfinder (Pielot and Boll, 2010) are two examples with a direct mapping that can guide the users with tactile sense. Both are belts with vibrators that indicate directions and deviations from the path. Another variant is the spatial tactile method (Yatani et al., 2012) that can offer high-level information about the distance and direction towards the destination by maintaining the spatial relationships between these points. Research by Dim and Ren (2017) to identify suitable body parts for vibration devices indicates that fingers, wrist, ear, neck, and feet had the highest perceptibility and user preferences. The study also suggests that the feet are unsuitable for vibration instruction in walking navigation.

Some assistive navigation devices use kinesthetic instruction modality to provide directional cues. Amemiya and Sugiyama (2010) developed a haptic direction indicator that delivers directional information in real-time through kinesthetic cues. The indicator uses a pseudo-attraction force technique to generate a force sensation.

Multimodal navigation systems hold the potential to enhance the accessibility of users with visual impairments. Several multimodal navigation systems have been proposed to assist users. Researchers also describe that navigation systems proposed for people with visual impairments could not be effective if the inputs and outputs depend upon only a single mode of interaction (Brock et al., 2015). According to Bainbridge (2004), a multimodal design can help improve the robustness of a navigation assistant system. As a result, when one of the modalities is unavailable or fails, a different one can be used (Grifoni, 2009). For example, vibratory instruction may be more effective than audio instruction in noisy navigation environments. In lieu, audio instruction modality could be an appropriate choice if the user wants to get more details about the environment, such as bus stops, traffic signs, landmarks, etc. (Kuriakose et al., 2020).

Different studies have reported on studying people's modality preferences using surveys and interviews (Quinones et al., 2011; Arditi and Tian, 2013; Ahmetovic et al., 2019; Ponchillia et al., 2020). This research aims to provide constructive findings that can be applied to the modality part of navigation assistant systems. Specifically, the main research question of this study is, *What type of information and mode of instruction would people with visual impairments prefer to have from a navigation assistant system?* To answer the research question, five hypotheses were formulated. The hypotheses were verified through experimentation and by evaluating participants' performance, experiences, and preferences. People with visual impairments and blindfolded participants with 20/20 vision participated in this study. The main contributions of this study are listed as follows:

1. A well-designed framework for a systematic experiment, execution, quantitative and qualitative evaluation, and analysis based on a survey and an interview process to study the modality preferences of people with visual impairments.
2. Importantly, the findings from the study on the effectiveness of two types of instructions, short and descriptive (which is not being explored by the research community), and two modalities, unimodal audio and multimodal audio with vibrations that could give new valuable knowledge in designing and developing future navigation assistant systems for people with visual impairments.

The paper is structured in the following manner. In Section 2, related works are presented. Section 3 details the study's methods, including study design, evaluation methods, and ethical analysis. Section 4 provides an overview of the experiment, including its setup and procedure. Quantitative and qualitative results are presented in Section 5. Section 6 delves into the discussions, while the conclusion is presented in Section 7.

2. Related works

Research conducted by Quinones et al. (2011) suggests that users need information regarding landmarks, the user's current location, roadworks, street names, and directions. Moreover, (Golledge et al., 1996) argues that when any form of environmental information is supplied to the users, words used to describe spatial configurations such as 'near,' 'around,' or 'a grey statue' may be ambiguous or inappropriate when describing environments to people with visual impairments. Sánchez and de la Torre (2010) suggested an alternative approach of using the clock system to inform the user of specific directions to reach a destination. Assuming that the user is facing 12:00, if we want the user to move to the right, we say, "Go to 3:00"; to go left, instruct the user, "Go to 9:00"; and to go backward, we say, "Go to 6:00". However, Nakajima and Haruyama (2013) and Miao et al. (2011) contradicted the clock approach with their results. They found that most users could not handle "clock hand" systems.

A survey by Arditi and Tian (2013) with ten users with visual impairments found speech as the preferred communication medium for navigating the environment using a camera-based assistive device. Another interesting finding was that the participants preferred to control the auditory display by querying the system rather than interacting via a menu or receiving a stream of continuous speech.

Ahmetovic et al. (2019) conducted an experiment with 13 participants with visual impairments to explore how the desirability of messages provided during assisted navigation varies based on navigation preferences and the expertise of users. They used NavCog, a navigation assistance tool that guides the users with turn-by-turn instructions enriched with contextual information about the environment. In the experiment, the participants were guided through two routes, one without prior knowledge and one previously studied and traversed. During navigation and follow-up interviews, the authors uncovered that participants have diverse needs for navigation instructions based on their abilities and preferences. A limitation of the study was that it seemed to overlook the effect of new routes on users' preferences.

Strothotte (1995) reported that users prefer information to be provided using synthetic speech rather than non-verbal audio or vibratory cues. In support of that, recent studies have also shown users preferring synthetic speech for navigation (Plikynas et al., 2020; Budrionis et al., 2022). With advances in natural language processing and text-to-speech conversion, synthetic voices sound more natural (Nguyen et al., 2023; Triantafyllopoulos et al., 2023). Loomis et al. (2005) compared non-verbal audio, synthesized speech, and/or vibrations initiated either from the hand's position or the torso. Auditory cues enabled participants to complete the tasks faster, and they generally liked spoken information about the distance to the next waypoint. On the other hand, some studies (Holland et al., 2002) also reported that users could use non-verbal sounds to portray information regarding the direction, distance, and location of landmarks concerning the user. Concurring with these findings, the studies done by Gaver (1991) reported that the use of beeps was less distracting than speech and easier to distinguish from environmental sounds. This method used the concept of waypoints to guide users, with beeps increasing in frequency as they were approached.

Lewis et al. (2015) presented a study to determine user requirements for navigation assistant systems. The study was done in two phases. The first phase was a focus group study where the results indicated that users require systems to provide information about their surroundings and progress. The next phase was an experiment establishing the best way to guide users between locations. They found that a preferred method of guiding users during navigation was a notification when they were both on and off track. Nonetheless, performance was best when only provided with the off-track notification. The results suggest that users should have control over the cues provided, and these cues should complement environmental cues rather than remove them. The study's main limitations mentioned in the article are demographic inconsistency and the participants' age who participated in the focus group study.

The study reported by Ponchillia et al. (2020) focused on identifying and figuring out the importance of preferences regarding what they found to be the most important type of information when navigating an indoor environment. The results showed that finding recognizable interest points and knowing their location is essential information. In addition, participants were asked to convey their preferred methods of receiving that information. The results showed that multimodal output through audio cues and vibrations was the most preferred.

The mental maps for visually impaired users proposed in Jacobson (1998) describe that audio clues denoting location, such as the sound of traffic or people, can help the user understand spatial configurations. However, the mapping of alternative senses such as hearing, touch, and smell could only sometimes be accurate since they do not have any direct mapping between various obstacles in the navigation environment and the user. Therefore, there is a need for research to understand favorable instruction modalities for designing navigation systems for the visually impaired.

Previous research shows various ways to develop impactful navigational systems for people with visual impairments. For example, different ways of using both unimodal and multimodal instructions from the environment can effectively improve how people with visual impairments navigate their daily lives (Kuriakose et al., 2020). However, to the extent of the author's knowledge and also from other studies (Real and Araujo, 2019), no solution is optimal in terms of adequate functionality, accuracy, and performance. In terms of past research, there are also evident limitations. Previous studies focused on testing and evaluating existing navigation systems or identifying user preferences through surveys. Thus, a research gap is identified in finding user preferences through a properly designed experiment followed by quantitative and qualitative analysis regarding relevant elements in navigation assistant systems, such as modality and instruction types.

Furthermore, many current research studies only experimented with blindfolded participants, making the findings much less reliable. Because not only does it put the blindfolded sighted participants in the position of working in a way foreign and unsettling, but being blind in this way is different from having no sight at all. As found in previous studies, multimodal output, mainly through audio instructions and vibrations, provides individuals with information about the environment (Ponchillia et al., 2020). However, no previous studies explored whether participants prefer to receive long descriptive or short and precise instructions or in combination with vibrations when navigating. This work aims at exploring this further through extensive experimentation with the users.

3. Methods

In the following sections, we describe the study design and various evaluation methods we used for this research, which include a performance evaluation experiment, a participant preference survey, and a participant experience interview. The section ends with risk analysis and ethical issue considerations.

3.1. Study design

This section describes the types and modality of instructions we considered in this study, the environment selection, and participant recruitment.

3.1.1. Type and modality of instructions

This study focuses on the effectiveness of two types of instructions: step-by-step and long descriptive, and two output modalities: audio and vibration, which are the commonly used instructions types and modalities in navigation assistants for people with visual impairments.

- *Short step-by-step instructions* are simple and direct instructions without any details. The participant receive only 'directions and turns' guidance during the navigation. An example of a step-by-step instruction is "turn left".
- *Long descriptive instructions* are instructions with complete sentences, like how a person gives descriptive instructions to another person. It also included detailed descriptions of a navigation environment, such as landmarks and obstacles, so the participant gets a more thorough mental picture of the surroundings. An example of descriptive instruction is, "One meter to your left, there is a trash can".

Many of the audio-based navigation assistant systems proposed in the literature use instructions similar to short step-by-step instructions (Kuriakose et al., 2022; Real and Araujo, 2019). However, based on our interaction with people with visual impairments during the initial experiment design phase, we understood that many users prefer to learn more about the environment while navigating. Hence, they prefer the navigation assistant to provide more detailed instructions. As a result of technological advancements, it is now possible to capture more information about the navigation environment, which can be delivered to users for a seamless navigation experience. To the best of the authors knowledge, no previous study has explored whether participants prefer long descriptive or short step-by-step instructions. Because of these reasons, we decided to include two types of instructions: short and long, in the experiment to gain knowledge about user preferences in the navigation context.

The instructions are given to the participant using two different modes. Audio mode is used in the case of unimodal, while audio and vibration are used for multimodal instruction. Audio modality was a common variable in all experiments after careful analysis. Researchers have different results on using audio modality as the output instruction mode in navigation experiments for people with visual impairments. For instance, Kuriakose et al. (2022) comments that audio instructions through headsets/headphones might block out ambient sounds and environmental cues, which people with visual impairments rely heavily on during their daily navigation. This could also potentially endanger people due to unforeseen risks they cannot detect. However, some studies favored the selection of audio modality as a choice (Golledge et al., 2004).

Moreover, one may insinuate that audio is preferable to provide an output if handled correctly. Therefore, when incorporating audio output into navigation systems for people with visual impairments, it is vital to consider the possible limitations, as these will affect the users interaction with the system and the surrounding environment. In addition, we need to test whether audio as a modality gives participants enough information and how the participants respond to short and

Table 1
Demographic information of blind/visually impaired participants. The table describes Participant ID, Age group: (20–29, 30–39, 40–49, 50–59), Gender: (Male, Female, Others), Level of visual impairment: (Moderate, Severe, Blindness), Age of onset (the age at which blindness happened) and usual navigation aids used by the participants.

ID.	Age group	Gender	Level of visual impairment	Age of onset	Usual navigation aids
22	40–49	F	Blindness	0	White cane, Guide dog, Personal assistant
23	50–59	F	Blindness	3	White cane, Personal assistant
24	40–49	F	Severe	5	White cane, Personal assistant
25	40–49	M	Severe	12	White cane
26	30–39	F	Severe	7	White cane, Personal assistant
27	20–29	F	Moderate	10	White cane
28	30–39	M	Blindness	0	White cane

long audio descriptions while navigating. Furthermore, vibrations can be considered one of the most used output modalities in navigation assistant systems for people with visual impairments to receive information in addition to audio (Cosgun et al., 2014). All these factors led us to include audio modality as one common modality in our study and vibrations in the multimodality combination with audio.

3.1.2. Environment

Some of the previous studies on navigation aids have been conducted both indoors (Lewis et al., 2015; Liu et al., 2021; Ahmetovic et al., 2019) and outdoors in more natural environments (Magnusson et al., 2010; José et al., 2011). To gain control over situational factors, it is beneficial to experiment in a controlled laboratory environment where the environment and the procedure are the same for all participants, regardless of when it is carried out (Tan et al., 2016). A controlled environment provides a stronger causal claim (Lazar et al., 2017). On the other hand, a field usability test is used to test the product in the environment in which it is to be used. Here, the researchers may also allow for additional aids during the test. For instance, in the experiments conducted by Lewis et al. (2015), Ahmetovic et al. (2019), participants could use additional navigation aids during testing. Moreover, in contrast to laboratory testing, the variables cannot be fully controlled during field testing but can be used in further analysis. Considering all these cases, we decided to conduct our experiments in an indoor controlled environment where participants cannot depend on any other assistive navigation tool such as a cane. In this way, we could obtain unbiased results from our experiment.

3.1.3. Participants

We tried to recruit participants with visual impairments (VI) through channels such as Blindeforbundet Oslo (Norwegian Association of the Blind) and Facebook groups dedicated to people with visual impairments in Oslo, Norway. However, as it was challenging to find enough participants, the problem was further compounded by COVID-19 pandemic restrictions at the time of the study. Hence we decided to conduct experiments with participants with visual impairment and 20/20 vision, and the latter blindfolded so that we have a good number of participants. We recruited seven participants who were visually impaired or blind and twenty-one with 20/20 vision for the experiment. Due to the small sample size of only seven VI participants, statistically robust conclusions cannot be drawn. Hence, we combined data from blindfolded participants with 20/20 vision with VI participants to compensate. Several studies have applied such a mixed participant sample by including blindfolded sighted participants (Barontini et al., 2020; Li et al., 2018; Apostolopoulos et al., 2014) and a study by Li et al. (2018) suggests that data collected from blindfolded participants could also provide valuable information.

The participants with a 20/20 vision were recruited from acquaintances not involved in any part of the experiment design. The demographics of seven participants with their level of visual impairment/blindness and their acquittance with conventional aids are given in Table 1. Participants numbers 1 to 21 were blindfolded participants. Among the total 28 participants, 19 were females, and 9 were males, with an age range between 20 and 58. All the participants had normal hearing and were adults.

Table 2
Different types of errors and their meaning.

Type of error	Description
Collision or a close encounter with an obstacle	The participant walked into an obstacle or the wall; alternatively, if we stopped them because they were on a possible collision course with an obstacle or wall.
Turning the wrong way	The participant turned less or more than 90 degrees or not according to the given clock instruction. Thus resulting in a deviation from the given path.
Walking in the wrong direction	The participant did not walk in the instructed direction or deviated from the path so much that they had to be stopped.
Falling	The participant fell during navigation.
Misinterpretation of the instructions	The participant did not act according to the given instruction. For example, if the participant turned left when instructed to turn right.

3.2. Evaluation

The study evaluates different combinations of types and modalities of instructions from three perspectives: quantitative performance evaluation, user preferences, and experiences. Performance is evaluated quantitatively through experimentation using the time the participants take to complete a path (source to destination) and the number of errors that occurred while navigating. Participant preferences are also evaluated quantitatively via a survey. Furthermore, participant experiences are analyzed qualitatively through interviews. Surveys and interviews were conducted on the same day as the experiment to ensure that information was not missed due to participants forgetting important impressions, thoughts, and experiences. The following sections describe the details of each evaluation method.

3.2.1. Performance evaluation experiment

The performance evaluation experiment employed a one-way within-subjects design. The independent variable was *navigation instruction mode* in four different trials: step-by-step instructions (audio only), descriptive instructions (audio only), short step-by-step instructions (multimodal), and long descriptive instructions (multimodal). The two metrics, time and error, used to evaluate the performance are defined as follows.

- *Time*: The amount of time, measured in minutes, a participant takes to navigate a track from start to finish.
- *Errors*: An error is counted if the observer has to relocate the participant back to the original path to complete the track. It is not counted as an error if they deviate from the path but still walk the correct way and do not face an obstacle. An error is also counted if the participant hits an obstacle. Each error was incremented from the base value of 0. Table 2 shows different types of errors, and Fig. 1 shows how deviations from the path are considered errors or not.

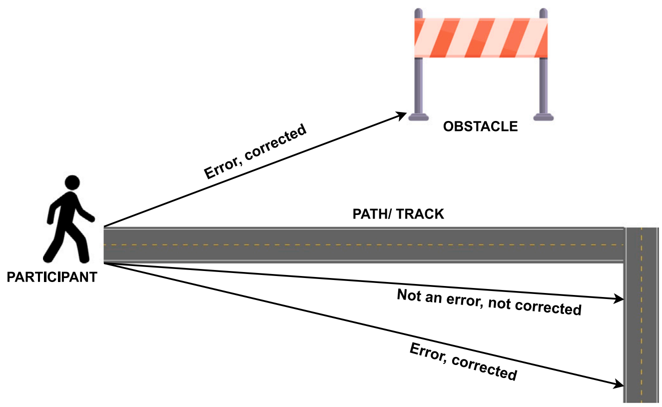

Fig. 1. The illustration shows how a small deviation from the path was accepted, as it is not natural to walk in a completely straight line. This was consequently not counted as an error, nor was it corrected. However, a larger deviation that resulted in a collision course with an obstacle was counted as an error, and the participant needed to be corrected to the prescribed path.

Three hypotheses are defined for quantitative analysis and statistical tests of the results from the performance evaluation experiment.

Hypothesis 1. Participants take less time to navigate if the navigation guidance instructions are short step-by-step instructions than if they receive detailed descriptive instructions.

Hypothesis 2. Participants will make fewer errors while navigating if they receive detailed descriptive instructions about the environment.

Hypothesis 3. Participants will make fewer errors while navigating if they receive multimodal (audio and vibrations together) instructions compared to unimodal audio instructions.

These hypotheses are checked for their validity through statistical analysis and tests. Before using the statistical test, the normality assumption was evaluated using the Shapiro–Wilk test. One-way ANOVA was used to examine group differences, whereas if there were only two groups, then a t-test was adopted.

Since the data were collected from individuals with visual impairments and 20/20 vision, two analyses were conducted to investigate whether a significant difference in performance could be determined. This analysis is to understand if there is any significant difference between participants with visual impairments and participants with 20/20 vision in time and errors while navigating. The meantime and errors (separately) of all trials, regardless of the instruction mode used, worked as the dependent variable. The independent variable was the visual status in two levels: 20/20 vision and visual impairment.

Before the experiment, a pilot test was conducted to determine if the experiment was feasible and could be adapted to a larger scale. The study evaluated the chosen experimental design, randomization, assessment procedures, guidelines, surveys, and technical procedures. The pilot test helped make the navigation track and obstacles safer for the participants and the survey questionnaire simpler and more understandable.

3.2.2. User preference survey

The survey conducted on participant preference consists of two-part questionnaires. The questionnaire was designed to explore how participants experienced the different navigation instruction modes to identify their preferences. The first part of the questionnaire consists of the following seven statements inquiring about the participant's experience of the different navigation instruction modes as experienced in the experiment.

1. I would prefer to use this instruction mode.
2. The instructions were complicated to follow while navigating.
3. It was difficult for me to follow the instructions when there was external noise.
4. I feel the system, as it appeared in the experiment, seems reliable.
5. I feel the system, as it appeared in the experiment, keeps me safe while navigating.
6. I feel like the instructions were missing some important information.
7. I feel like there was given too much information during the navigation.

A five-point Likert scale was used for the survey (with values ranging from strongly disagree, disagree, neutral, agree, and strongly agree). The questionnaire was repeated after each experiment trial, resulting in four repetitions per participant.

The second part of the questionnaire was asked post-experiment. Here, the participants were asked to rank (from 1 to 4, where 1 is the most preferred) the four instruction modes (step-by-step description, descriptive information, step-by-step description with vibrations, and descriptive information with vibrations) based on their preferences.

Previous research has found that even the smallest changes to an attitude question had unusually large effects on the answers (Groves et al., 2011). Therefore, the questions were carefully worded and asked the same way for each participant. Two hypotheses are defined for quantitative analysis and statistical tests on participant preference.

Hypothesis 4: Participants prefer to hear details about the environment and obstacles during navigation.

Hypothesis 5: Participants prefer multimodal instructions rather than unimodal audio instructions.

3.2.3. User experience interview

User experience data is collected through interviews. Interview questions were open-ended questions intended to learn about the participant's choice of the type and mode of instruction and the reasons behind their choice. Collected data are analyzed qualitatively using a method of categorical division (Johannessen et al., 2010). Data is compiled into one document and divided into different topics, with each participant's number as a subcategory. Further, all the data is marked using color-coding to differentiate between the different opinions of the participants. This is placed into five categories; vibration, descriptions, turning and rotating, personalized information, and external effects. Then each category is looked at thoroughly, counting how many have the same opinion about the various topics and extracted quotes that has meaningful information to the results. The analysis is done based on a meaningful division of topics mentioned and categorizes the common features of the participants' opinions. Collected data is viewed objectively, and interpretation of the data to give specific meaning to what a participant said is avoided. Therefore, quotes are used when participants specifically state his/her view.

3.3. Risk analysis and ethical considerations

We did an extensive risk assessment and prepared a mitigation plan before experimenting to reduce risks and uncertainty during the research. No private or personal information was collected, stored, or analyzed. To address ethical issues, we followed guidelines from General Data Protection Regulation (GDPR)[1]. Following the Norwegian Agency for Shared Services in Education and Research (Sikt)[2] guidelines, a standard consent form was used to get consent from the

[1] https://gdpr.eu/
[2] https://sikt.no/en/home

participants, and a proper data management plan was put in place for its safety and security.

Two different consent forms were used for participants with visual impairments and 20/20 vision. The consent forms included the same information, except that participants with 20/20 vision were informed about being blindfolded and instructed not to remove it within the experiment room. In contrast, the participants with visual impairments were asked not to use any other navigation aids during the experiment. Moreover, participants were informed of their right to withdraw from the experiments without giving any reasons.

4. Experiments

Four experiments were conducted for the four combinations of the two types of instructions using two different modalities. Four different tracks (see Fig. 2) were created for these experiments. The experimental setup and the procedure are described in the following sections.

4.1. Experimental setup

To conduct the experiment effectively and efficiently, various roles were defined for those involved. The roles include *Controller 1*, *Controller 2*, *Observer 1*, *Observer 2*, and *Interviewer*. *Controller 1* was responsible for providing audio instructions to participants and monitoring whether the participant was on track. *Controller 2* was responsible for providing vibrations to participants and monitoring whether the participant was on track. *Observer 1* recorded the time the participant took to complete each experiment trial, kept track of the number of errors and the time asked for assistance, and conducted additional observations during trials. *Observer 2* was concerned with observing and correcting the participant if they made an error. The *interviewer* was responsible for the consent form and survey, verifying that the participant answered questions after each trial before moving on to the next track. In the following sections, we describe the environment and tracks created to conduct the experiments and the devices and software applications used to give participants different types and modalities of instructions.

4.1.1. Environment and tracks

The experiment was conducted indoors in a 46 m^2 size room. The room was not disturbed by external interferences or external noise except simulated background noises. These background noises were played with two speakers at opposite ends of the room to simulate a natural navigation environment. One of the speakers played traffic sounds, whereas the other was used to play nature sounds. A first aid kit was also kept in the experiment location for the unlikely event of physical injury.

Four tracks were created for the four experiments. The first two tracks (tracks 1 and 2) were used to test unimodal audio-based short step-by-step and long descriptive instructions. The other two tracks (tracks 3 and 4) were created for multimodal (audio+vibration) step-by-step and descriptive instructions. Fig. 2 shows the layout of navigation tracks used for four experiments.

Each track was exactly 40 m long. Obstacles of various types and sizes were placed in the tracks. Different things, such as big cardboard boxes, trolly, styrofoam, baskets, and plastic boxes, were used as obstacles and placed in the tracks. All tracks used the same obstacles but were placed in different directions corresponding to four experiments. A photo of the experiment room with tracks and obstacles is shown in Fig. 3.

Table 3
Vibration patterns used for the multimodal instruction modes.

Button	Vibration	Vibration length
Forward	Three short vibrations	250 ms × 3[a]
Stop	One long vibration	900 ms

[a]Delay between consecutive vibrations: 500 ms.

4.1.2. Devices and software applications

The experiment used two headphones to ensure the participant and the controllers could hear the audio instructions. The participants were equipped with a Bose QuietComfort 35 headset, whereas the *controllers* shared a pair of AirPods. This ensured that the instructions received on the participant side were in sync with the *controllers* side.

A desktop application was developed to control the short and long voice instructions (see Fig. 4). The desktop application consists of a direction-controlling panel, a clock panel corresponding to the angle at which the participant needs to turn, and a set of short and long instructions for different buttons (This is different for both tracks). Audio instructions were played in synthesized American English, resembling a female voice. The speech rate was standard and never adjusted in any of the trials or for any of the participants. The desktop application was connected to both wireless headsets via Bluetooth.

The participant was also equipped with a smartwatch to receive vibratory instructions. The smartwatch used was the Fossil Gen 5E. Vibrations were controlled from an android application (see Fig. 5) at the *controllers* side. It was not possible to control the frequency and amplitude of vibrations of the smartwatch. Hence we used the smartwatch with the system-level default settings. Two vibration patterns were used, signaling different instructions (see Table 3). The smartwatch was connected to the Android application using WiFi. The synchronization between audio and vibration instructions was ensured by the two *controllers*. The two controllers followed the same navigation-step script for the four different tracks while seated next to each other. Thus, both controllers coordinated simultaneously to give directions to participants.

4.2. Experimental procedure

At first, the participant is brought to the experiment location. The first part of the study is in a separate room next to the experiment room. This can help to prevent sighted participants from seeing the tracks before the experiment. Before starting the experiment, the experiment details are shared with the participant. Participants provide oral consent to the consent declaration read to them. Then the participants with 20/20 vision are blindfolded using a pair of blacked-out ski goggles to ensure they never see the tracks before completing the experiment.

Upon entering the experiment room, the participant is given information regarding the instructions to familiarize themselves with the navigation modes. This includes a tutorial on how to respond based on the different audio and vibration navigation instructions and how far they should turn when asked to turn left or right. They are also informed they should *raise a hand* if they want any instruction repeated, etc. The participant is provided with headphones and a smartwatch to receive instructions. The *Contoller 1* also has headphones to ensure that the instructions received at the participant's end are the same. Then the *Observer 2* guides the participant to the starting point for their first track. *Observer 1* uses a stopwatch to measure the time the participant takes to complete the trials.

The Wizard of Oz method (Bella and Hanington, 2012) is used in each experiment trial to guide participants through the tracks. To even out the learning effect across tracks, the order of completion is randomized. The two *Controllers* provide navigation instructions, delivering prerecorded audio instructions through the headset and vibrations through the smartwatch to the participant. Instructions are based on pre-made scripts for each of the four tracks. The *Observer 1*

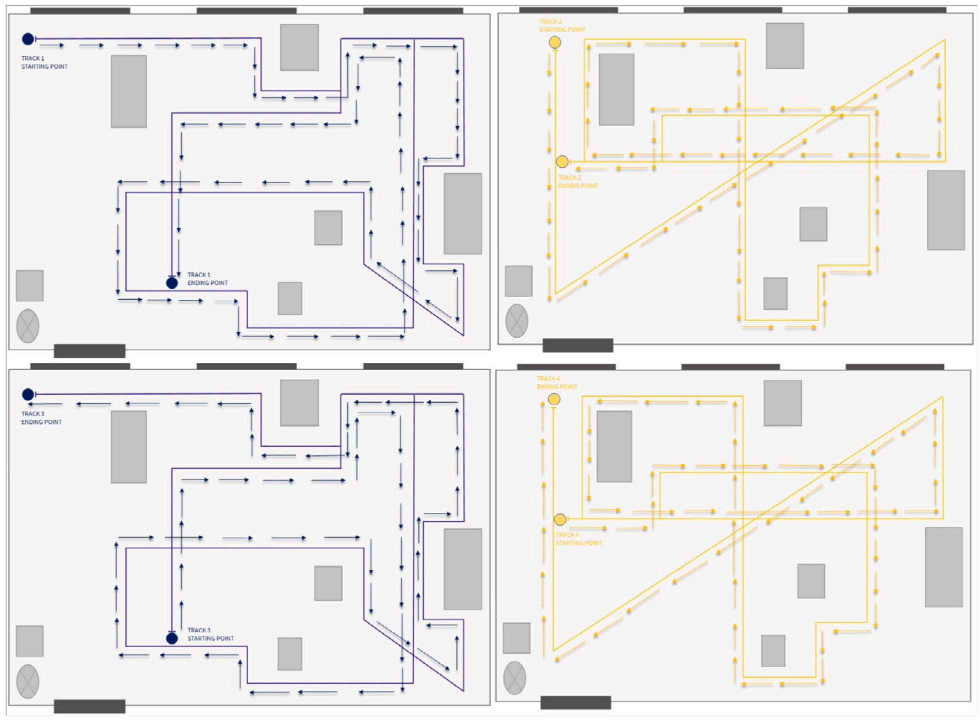

Fig. 2. Illustrations of the experiment room and the tracks. Tracks 1 and 3 have the same tracks but in reverse directions. Tracks 2 and 4 have the same tracks but in reverse directions. The starting point of a track becomes the endpoint for the other and vice versa. The arrow marks indicate the direction of navigation in respective paths.

Fig. 3. A photo of the room showing navigation tracks and obstacles. Two tracks were created (refer to Fig. 2). Tapes in two different colors (yellow and blue) were used to indicate the navigation path. The two tracks were made in both directions, resulting in four tracks.

records the time taken to complete each track and the number of errors made by the participant. The *Observer2* accompanies the participant to ensure no harm or injury occurs at any given time. Each participant must complete all four tracks, resulting in four different tracks.

After completing each track, the *Interviewer* asked the first part of the user preference survey to the participant about the current instruction mode. Afterward, user experience interviews were conducted. After completing all four trials, the second part of the user preference survey was conducted by asking the participant to rate all the instruction modes and asks to give their overall impressions of instruction methods. Each participant spends approximately one hour completing the whole procedure.

5. Results

This section presents the results in detail in three sections: performance, preference, and experience of the participants.

5.1. Performance

The statistical results after analyzing the time performance, errors made on each instruction mode, the between-group differences, and modality preference analysis through the survey are presented here.

5.1.1. Effect of navigation mode on time

A Shapiro–Wilk test found no indication of non-normality for either of the conditions: step-by-step (W = 0.93, p = 0.06), descriptive (W = 0.96, p = 0.39), step-by-step with vibrations (W = 0.95, p = 0.16) and descriptive with vibrations (W = 0.97, p = 0.66). The ANOVA could thus be applied and yielded a statistically significant effect, $F(3, 81) = 9.84$, $p < 0.001$, partial $\eta^2 = 0.27$. To further evaluate the difference between the four means, paired samples t-tests were used. The t-test found no significant difference in the time spent completing a track when step-by-step instructions were used compared with descriptive instructions, $t(27) = 0.46$, p = 0.65. Thus, the hypothesis that participants would take less time to navigate if instructions were short was not supported. However, further analysis indicates a significant difference in time spent completing a track with descriptive instructions compared to descriptive instructions with vibrations, $t(27) = -2.34$, p = 0.03, and in the time spent completing a track when step-by-step instructions are used in comparison to step-by-step instructions with vibrations, $t(27) = -3.76$, p = 0.001. For both cases, the meantime is lower for unimodal audio instructions, indicating that more time is generally spent navigating when vibrations are added to the instructions (see Fig. 6). However, interestingly, if we consider the participants with visual impairments alone, we can see that the mean time taken step-by-step (audio alone) is not significantly less than the descriptive instructions (audio alone). But on the other hand, when vibrations are included, the mean time taken with descriptive instructions is significantly less than the step-by-step.

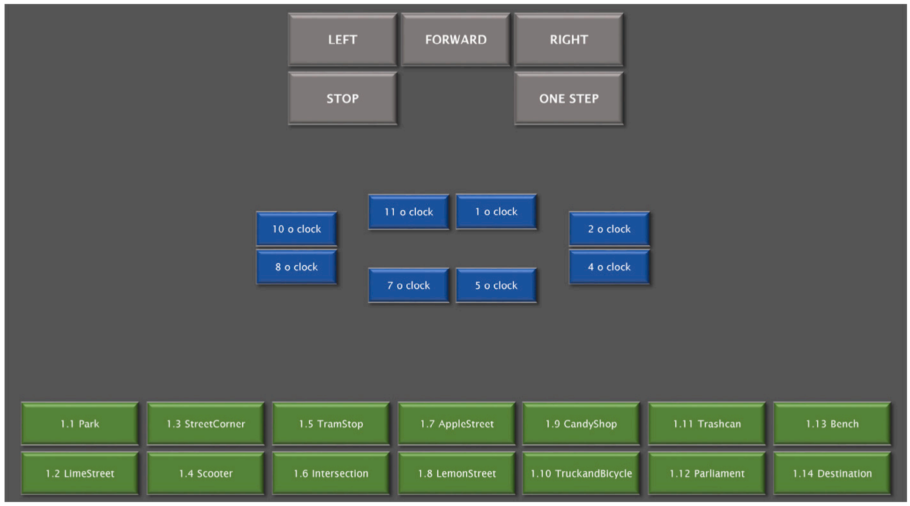

Fig. 4. The desktop application used to provide participants with detailed audio instructions for track 2. In the upper part of the screen, buttons controlling instructions related to turning, moving, and stopping can be seen. The buttons controlling the descriptive instructions are at the bottom of the screen.

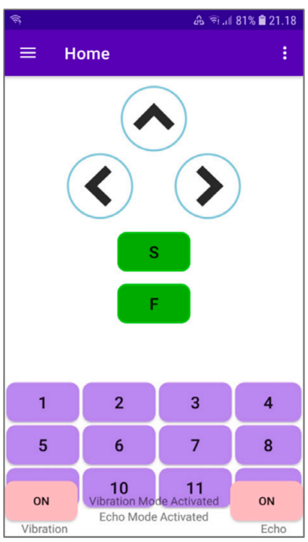

Fig. 5. Smartphone application used to control vibrations in a smartwatch. The **S** button corresponds to *stop* instruction, and the **F** button corresponds to *One step forward* instruction. The other buttons were not used in this experiment.

5.1.2. Effect of navigation mode on the number of errors made

A Shapiro–Wilk test found no indication of non-normality for either of the conditions: step-by-step (W = 0.97, p = 0.52), descriptive (W = 0.95, p = 0.20), step-by-step with vibrations (W = 0.93, p = 0.07), and descriptive with vibrations (W = 0.95, p = 0.18). ANOVA test identified no significant effect of instruction mode on the number of errors made, F(3, 81) = 1.17, p = 0.33, partial η^2 = 0.04. Looking at the descriptive statistics (see Fig. 7), it is clear that there is a difference in the mean number of errors for each track and that fewer errors are generally made when receiving detailed instructions (M = 5.57) compared to step-by-step instructions (M = 6.11). However, the difference is minimal (0.14) and may be due to chance. As such, the test result provides no evidence supporting the hypothesis that participants will make fewer errors while navigating if they get detailed information about the environment (detailed instructions) than if they receive short instructions (step-by-step). Furthermore, from Fig. 7, the mean number of errors is higher for the multimodal step-by-step (M = 6.64) and descriptive (M = 6.07) conditions than for the unimodal versions of the two instruction modes. Hence, there is also no support for the hypothesis that participants will make fewer errors while navigating if they get multimodal instructions compared to unimodal audio instructions. But in contrast, if we consider only participants with visual impairments (see Fig. 7), the results indicate that with multimodal instructions, the number of errors made by the participants is reduced when it is step-by-step (M = 4.29) and descriptive (M = 4.43) even though it is not statistically significant.

5.1.3. Between-group differences in time and error performances

To assess how participants with 20/20 vision and participants with visual impairments performed about the time taken to complete the tracks, an independent samples t-test was used. The t-test found a significant effect of visual status on-time performance, t(26) = 3.59 p = 0.001, where the participants with visual impairments (M = 4.67) generally completed the tracks in significantly less time than the 20/20 test group (M = 5.78).

The same tests were applied to the number of errors made by participants with 20/20 vision and with visual impairments. The independent samples t-test found a significant effect of visual status on the number of errors made, t(26) = 2.07, p = 0.049. On average, participants with 20/20 vision (M = 6.6, SD = 1.75) made significantly more errors than participants with visual impairments (M = 4.61, SD = 3.29). Consequently, these two between-measures analyses prove that participants with visual impairments spend less time and make fewer errors while navigating than participants with 20/20 vision.

5.2. Preferences

After completing all trials, participants rated all instruction modes from most preferred to least preferred. Fig. 8 presents the results from this question graphically. The diagram shows that the multimodal descriptive instructions were rated as the most preferred instruction mode by the largest number of participants (10 votes), followed by the unimodal audio descriptive instruction mode (7 votes). This implies a preference for the descriptive instruction modes, as was hypothesized. According to the ratings for least preferred instruction modes, the unimodal step-by-step mode received the largest number of votes (12), followed by multimodal descriptive instructions (8 votes) and unimodal descriptive instructions and multimodal step-by-step, with four votes. As such, it is apparent that there is a significant gap in participants' preferences, as the multimodal descriptive instruction mode both received a large number of votes in the most preferred and least preferred categories. Moreover, the graph shows that the participants with visual impairments also prefer descriptive information compared to short instructions. This is indeed true since, in the real world, people

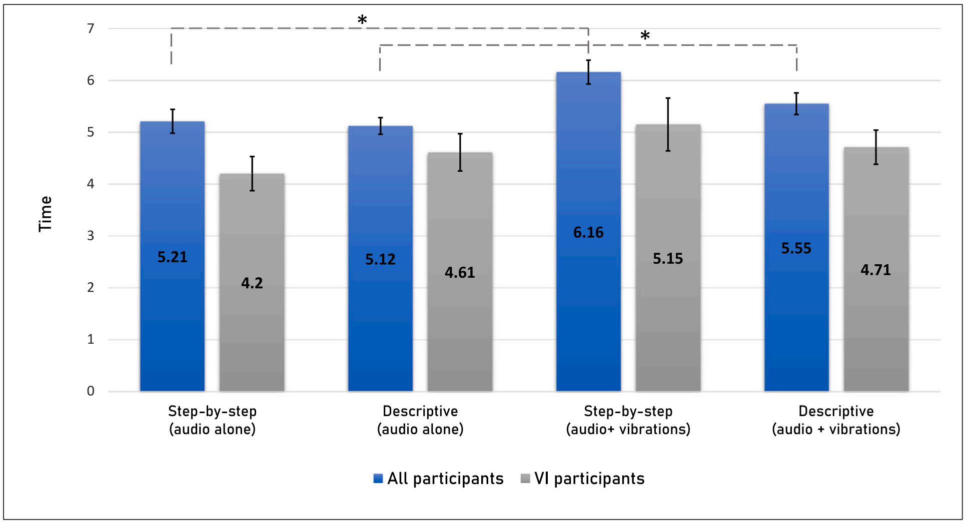

Fig. 6. Graphs showing the average and standard error of time taken (in minutes) in different navigation modes. The significant pairs are shown by * sign.

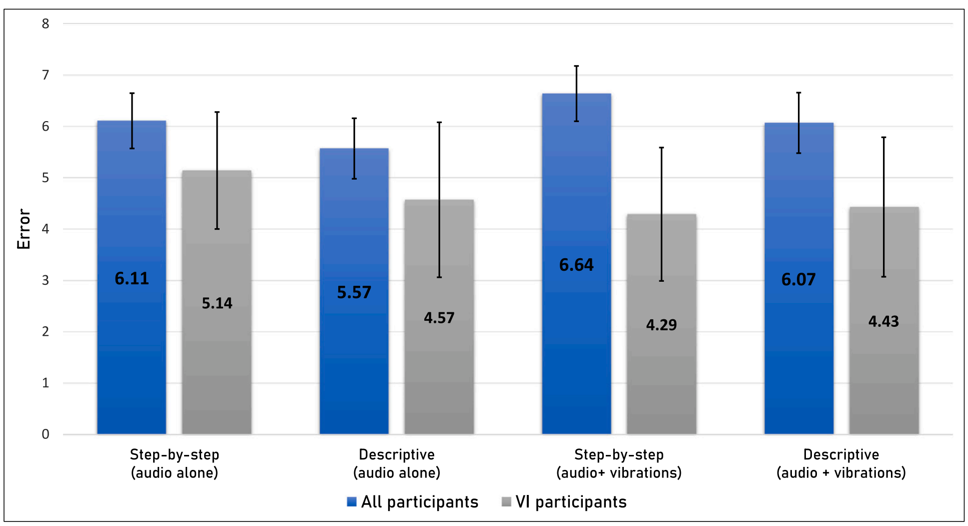

Fig. 7. Graphs showing the average and standard error of error counts in different navigation modes.

with visual impairments prefer to have more information to navigate safely in unknown environments. Another interesting observation is that unimodal descriptive instruction was rated as the most preferred instruction mode, and unimodal step-by-step was rated as the least preferred instruction mode by the participants with visual impairments.

The user preference survey is analyzed to test the hypothesis that participants prefer to hear details about the environment and obstacles during navigation. The mean scores from the scales completed after tracks 1 (step-by-step) and 2 (descriptive) were calculated separately and applied with a paired samples t-test. Before the analysis, the score of negatively worded items had to be reversed to ensure all variables were consistent.[3] In evaluating the assumption of normality, the Shapiro–Wilk test found no indication of non-normality for either condition (step-by-step, $W = 0.95$, $p = 0.19$, and descriptive, $W = 0.93$, $p = 0.41$). The test yielded no statistically significant effect, $t(27) = -0.39$, $p = 0.67$. In other words, the measured strength of participants' attitudes and perceptions of the descriptive instructions ($M = 3.99$, $SD = 0.64$) did not significantly differ from the short instructions ($M = 3.95$, $SD = 0.6$). The result contradicts the hypothesis that participants prefer to hear details about the environment and obstacles during navigation.

Similarly, to test the hypothesis that participants prefer multimodal output for receiving navigation instructions, as opposed to unimodal instructions, the mean score of the Likert scale questionnaire of tracks 2 (descriptive) and 4 (descriptive with vibrations) were compared, as well as track 1 (step-by-step) and 3 (step-by-step with vibrations). A Shapiro–Wilk test found no evidence of non-normality for either of the multimodal instruction modes (step-by-step with vibrations, $W = 0.94$, $p = 0.12$, and descriptive with vibrations, $W = 0.95$, $p = 0.25$). A paired samples t-test found no significant difference in the measured strength of participants' attitudes and perceptions of the instruction modes: descriptive ($M = 3.99$ $SD = 0.64$), descriptive with vibrations ($M = 3.96$ $SD = 0.58$), $t(27) = 0.44$, $p = 0.67$. However, the t-test did yield a significant result between the instruction modes step-by-step and step-by-step with vibrations, $t(27) = 2.14$, $p = 0.04$, where the measured strength of participants' attitudes and perceptions of the step-by-step instructions ($M = 3.95$, $SD = 0.6$) were significantly higher than the instruction mode step-by-step with vibrations ($M = 3.73$, $SD = 0.70$). Thus, the result also offers no evidence to support the hypothesis.

To draw a cohesive picture of the participants' preferences towards the different instruction modes, we used the *weighted scoring*

[3] https://guides.library.lincoln.ac.uk/c.php?g=110730&p=4656824

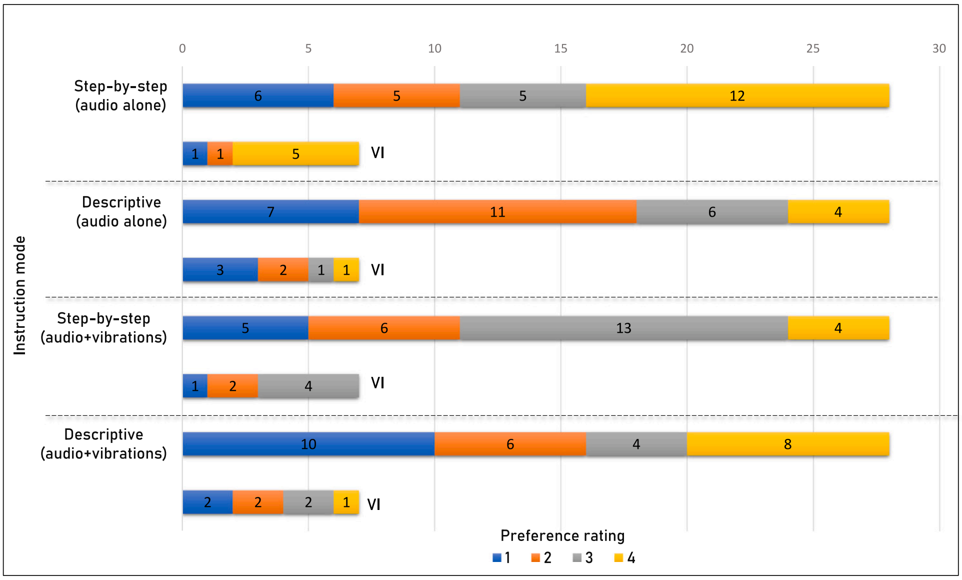

Fig. 8. Preference ratings of all participants for each instruction mode. Instruction modes are presented on the *y*-axis, and the preference ratings are on the *x*-axis. Instruction modes are rated from 1 (most preferred) to 4 (least preferred). The bars marked with 'VI' result from participants with visual impairments alone.

method (Jadhav and Sonar, 2009) to illustrate which instruction mode received the best overall rating. For this evaluation, we assigned a score of 1 to 4 to the different ratings in Fig. 8, where 4 points equaled the most preferred instruction mode, and 1 point equaled the least preferred instruction mode. The total score for each instruction mode could then be calculated. For instance, the most preferred instruction mode for descriptive instructions (audio alone) received seven votes ($7 \times 4 = 28$ points). Adding the remaining three placements ($11 \times 3 = 33$, $6 \times 2 = 12$, $4 \times 1 = 4$) leaves the instruction mode with a total score of *77 points* and the highest overall score of all the instruction modes. Similarly, descriptive got *61 points*, step-by-step received *20 points*, and step-by-step with vibrations received *12 points*.

Furthermore, we used the same procedure to calculate the score from the results obtained from the participants with visual impairments alone. Similar to the results of all participants, the descriptive instructions also received the highest score rating for participants with visual impairments, with *70 points*, descriptive received *59 points*, step-by-step got *25 points*, and step-by-step with vibrations got *14 points*.

5.3. Experiences

As mentioned in Section 3.2.3, the experiences collected from the participants through open-ended interviews were categorized into various groups. They are described as follows.

5.3.1. Vibrations

The qualitative data about multimodal navigation instruction showed that vibration was valued by some participants, allowing them to move around more quickly. Half of the participants mentioned that they enjoyed using vibrations through a smartwatch. Among them, five participants were with visual impairments.

Many participants who enjoyed vibrations stated that it was an essential asset in receiving information while walking, and it confirms the audio information they received. Some participants mentioned that vibration worked well in places where audio was insufficient, and vibration worked well as a second mode confirming the source of information. Various participants also suggested that vibration can be seen as an asset when faced with a lot of external noise. Few participants said that vibration was good for preventing misunderstandings.

> *"I felt like vibration was good to have, to have another indicator for when I was to move forward or stop that couldn't misinterpret as descriptive information".* (Participant 8)

Some participants also suggested that vibration could help eliminate unnecessary audio instructions. One participant mentioned that 'stop' and 'move forward' would work better if they were given through vibrations, not audio. A few participants mentioned that they felt they could move faster and with a better flow when using vibrations than when all information was given through audio instructions.

Among the 11 participants that did not enjoy using vibration as a method of instruction, most reasoned it by saying that it was too much information at once and that there was too much to pay attention to.

> *"I felt that the vibration was a bit too much, especially when there were voice instructions as well. I got a bit confused when there were many things to keep track of, such as vibration and sound".* (Participant 19)

Some participants expressed that vibrations would work better if they could come with adjustable patterns.

> *"I felt navigation was a bit excessive with both voice and vibration. I might have enjoyed it better if the vibration pattern could be adjustable".* (Participant 28)

5.3.2. Descriptions

20 of 28 participants (among them, six were participants with visual impairments) reported that they enjoyed long descriptions over short ones. Six participants, of which one was a participant with visual impairments, mentioned that they enjoyed short descriptions. Other participants specified that they enjoyed both.

There were various reasons why participants enjoyed the audio instructions with long descriptions. Participants said that long descriptions gave them better confidence when navigating, giving them a sense of security and a feeling of going in the right direction. It was also mentioned that the long descriptions gave them greater walking options. Many participants revealed that long descriptions provided a clearer image of what was happening around them. Another reason for preferring long descriptions was how the participants got to interact with the outside world, *"I enjoyed the descriptions because that's what this is about, to get a visual interpretation, which makes me able to, it gives me the freedom to choose, maybe I want to walk into that restaurant, or maybe I know there is a trash can in front of me, I needed that".* (Participant 24)

Among the five participants who did not enjoy short descriptions, they explained it was due to not knowing what was around them, making them feel unsafe.

"I didn't know exactly what's around me, so how safe I was, I didn't have any idea about the surroundings". (Participant 26)

But a few participants did not enjoy long descriptions and could not see the need for all that information. According to them, it was easier to follow the instructions when they were short. And they did not feel like they needed to know everything around them all the time.

5.3.3. Turning and rotating

Features like turning left/right, turning clockwise, and being corrected if they deviate from the path indicated mixed opinions from the participants' experiences. Some participants wished the system would correct them if they had turned incorrectly. One participant said that, in a real-life setting, it would be easier to use the application with the help of being corrected. Many participants also specified that understanding how much to turn took a lot of work.

"You have no one to correct you, so you won't know when you are about to turn to the right, and you believe that you have done it correctly, but you might have turned too much".(Participant 21)

There were also mixed feelings about the instruction about turning clockwise. One participant believed that turning by the clock worked better than being told to turn left and right. Some participants, however, had trouble understanding how to turn clockwise.

"It was difficult to understand when I had to turn by the clock. I can understand the direction, but not exactly how much I have to turn". (Participant 19)

Few participants had suggestions to make turning more intuitive. For example, one suggested only using the top half of the clock.

"It is intuitive to follow the clock rotation from 9–3 (the different numbers in front of you), but it's more difficult to turn when it comes to the numbers behind you. I prefer to be told to turn left, then get a clock rotation for more accurate turning". (Participant 25)

Participants commented that turning correctly was heavily influenced by their opinions about how much turning was necessary.

"You must practice what they mean when they say turn left and turn right, and practice what that means for you". (Participant 18).

5.3.4. Personalized information

Eleven of which six were participants with visual impairments voiced a desire to personalize the information they were given during navigation. Many participants wanted the possibility to customize the information given based on what situation they were in, deciding what type and how much information they were given.

"I am very dependent on the context regarding how much information I want to receive and my interests. A trashcan might be useful sometimes, but other times it can be very unimportant". (Participant 28)

A few participants mentioned that detailed information about tram/bus stops when they are coming, their time of arrival, etc., could also be helpful. Some participants mentioned that they would like to customize what information is given through audio instructions and what is given through vibrations.

"It would be nice if I could choose what information I want through audio and vibration, make adjustments and choices to vibration patterns myself and vary the descriptions and tempo of the voice and vibrations, and decide the purpose of the vibration". (Participant 22)

One of the participants said that it would be beneficial to have two different voices give different types of audio information:

"Even though it's nice to listen to what is happening around me and get a picture of what's around, it becomes confusing when mixed up with the walking instructions. I had to stop and listen when the voice came if I was getting instructions or a description. I think it would work better if two voices were used, one for instructions and one for descriptions". (Participant 7)

5.3.5. Other observations

Twenty participants (six with visual impairments) expressed concerns about using a technology-aided device. According to them, users likely became less social and detached from the outside world.

"This could also take my attention away from other things, like things around me, if someone says hi or they need help, but because I'm so focused on the system, vibration will be a benefit in this situation". (Participant 24)

Many participants mentioned that faulty equipment, or other technical difficulties, could lead to unwanted or dangerous situations. Vibrations can also be misleading if a smartwatch is also connected to a phone and when they receive texts or calls.

"If we talk about the technology if it fails, you can quickly become very helpless if you suddenly have to walk around without it because you haven't trained your navigation". (Participant 26)

6. Discussion

Here, we discuss the validity of hypotheses from the results, limitations, and future perspectives.

6.1. Validity of hypotheses

The statistical analysis of the results from the experiments found no significant difference in the time spent completing a track when short step-by-step instructions were used compared to long descriptive instructions. The results may be a consequence of the setup of the tracks. Both step-by-step instruction (audio) and step-by-step instruction (multimodal) modes were completed on one track, and both descriptive instruction modes were completed on the other. Both those tracks were the same length long, but the track for descriptive instructions had fewer turns than the step-by-step track to accommodate the instruction modes. For instance, the track used for descriptive instructions needed longer stretches, as more information would be provided during navigation. One would assume individuals would spend longer navigating when instructions are descriptive, as more time is spent listening to instructions rather than just walking. On the other hand, the additional number of turns of the step-by-step tracks may have increased the possibility of turning incorrectly and thus resulting in more time spent on corrections. Therefore, the tracks with step-by-step navigation may have been more time-consuming for the participants than first assumed. Consequently, the nature of the tracks may have eliminated potential time differences for the different navigation modes as they may have occurred in a more natural environment. Although there was no significant difference in the time spent navigating between the step-by-step and descriptive instructions, the analysis found that participants spent more time navigating by multimodal instructions than audio-only.

The results showed no significant effect of instruction mode on the number of errors the participants made during navigation. As mentioned before, more turns during the step-by-step tracks may have increased the possibility of correcting the participants on their turning. Before starting the experiment, the participants were told to turn by what they perceived as 90 degrees when told to turn left/right. As this is subjective, it was observed that most participants failed to turn correctly on several occasions. Due to the increased number of turns in the step-by-step tracks, the participants had a more frequent possibility of angling themselves incorrectly than when they walked the tracks with descriptive instructions, with fewer turns. On the other hand, the tracks with descriptive instructions had longer stretches than the step-by-step tracks, which could increase the chances for the participants to deviate from the path. As such, this could explain the non-significant result.

The study also found no significant difference in the number of errors between multimodal and audio-based instructions. As the synchronization between audio and vibrations was sometimes lacking, participants might have received conflicting instructions, thus leading to the occurrence of more errors than if the two modalities had been properly synchronized.

The study showed attitudes and perceptions of the participants on the descriptive instructions did not significantly differ from those of the step-by-step instructions. The participants stated that they could see the benefits of both types of navigation instructions but that this was contextually based. It was desirable to have long descriptions in some situations, for example, when traveling on vacation, since the descriptive instruction will allow for spontaneity, which is experienced as difficult according to several participants. Participants argued that step-by-step descriptions could be practical in places where they are already familiar with the environment, where the purpose is to navigate between places quickly and efficiently, such as when going to work or university.

Similarly, no significant difference was found in the attitudes and perceptions of participants between audio-alone and multimodal instruction modes. One of the reasons for this could be that the vibration was not synchronized with the voice instructions a few times, as it comes with a delay after the voice instruction. Some participants pointed this out, and it may have affected their way of responding. Moreover, some participants did not show an inclination towards the vibration regardless of synchronization issues that happened a few times, as they perceived it as overwhelming information in combination with audio. However, many participants, especially the visually impaired participants, said vibrations would be a practical addition to the audio instructions in outdoor situations when accompanying other people and in conversations, where vibration could serve as an additional source of information.

Participants with visual impairments spent less time navigating than participants with 20/20 vision. The reason may be that participants with visual impairments have more experience navigating without sight and therefore completed the trials. Some participants with visual impairments expressed that they walked slower during the experiment to our benefit, as they understood they were inside a smaller room. However, participants with visual impairments still walked quite fast, making it challenging to ensure they got to experience each navigation instruction. It would have been a challenge if the participants with visual impairments had walked at their usual pace. Participants walking speed and how it may affect their navigation should be assessed in future experiments. Moreover, most participants ambled on their first trial but picked up the pace as more tracks were completed. The fact that the research team assured the participant that they would avert them from injury might have made the participants feel safer, making them increase their walking speed after getting used to the experimental conditions.

It is visible that blindfolded participants made significantly more errors than participants with visual impairments. This might be because participants with impaired vision are used to navigating without sight, while blindfolded participants are not. Moreover, the result that blindfolded participants spend more time and make more mistakes is not surprising, as it confirms the importance of being aware of the differences between these groups when conducting experiments, implying that blindfolded participants cannot replace participants with visual impairments.

6.2. Limitations

There were some limitations in this study. The experiment was conducted in a relatively small room of 46 m^2, with short distances between obstacles and turns. This may have affected the participants' experience with the navigation instructions, as they could only move a very short distance before receiving a new instruction. This was especially apparent in the tracks where descriptive instructions were used. Hence the experimental setting likely made the navigation experience less natural than if it had been performed in a larger area with uneven surfaces and natural obstacles.

Similarly, the number of errors registered might have differed if the experiment had been performed outdoors or in a larger area. It could have led to participants having more freedom to move without necessarily deviating from the track. However, experimenting in an outdoor environment would challenge the control over situational factors. As claimed by Patel (1995), control is essential in experimental research, and this is to eliminate external factors that can affect the results. Moreover, conducting the current experiment outdoors would not be an ideal solution. However, replicating the experiment in a larger indoor space with several levels could be beneficial to check if the results would be different in a more natural setting.

Having more participants with visual impairments would have strengthened the experiment results, as the findings would represent actual end-users of assistive navigation technology. Getting a good number of participants with visual impairments proved challenging, and COVID-19 restrictions at the time made it even more difficult. We understand that people with visual impairments have developed a range of workarounds for various situations. This includes utilizing alternative forms of information, such as echolocation, and having a clear awareness of the types of mistakes that can lead to disorientation or getting lost (Thaler and Goodale, 2016; Pasqualotto and Proulx, 2012). Sighted participants may not have the chance to develop these strategies Kolarik et al. (2014), Tinti et al. (2006). Hence the results interpreted from this study involving blindfolded participants should be viewed from this perspective. Moreover, the seven participants with visual impairments we managed to recruit can be considered satisfactory considering the circumstances (McDonald et al., 2006).

The most common mistakes/errors observed during the experiment were that participants frequently turned too much or too little when instructed to "turn left" or "turn right" or were asked to turn according to the analogous clock. As such, it may be necessary to consider whether these "errors" affect the different instruction modes or whether they result from personal perception. Consequently, it could be argued that the preferences identified through the survey are a more appropriate basis for determining the most effective instruction mode.

6.3. Future perspectives

Contrary to previous research reported in the literature, in our study, participants with visual impairments did not use their usual navigation aids, such as a white cane or guide dog. The participants were positive about experimenting without them and expressed curiosity to try different types of navigation aids. This observation could be helpful for future studies and potential new solutions regarding not being restricted by using such aids. Furthermore, bone-conduction headphones can be used instead of standard headphones. The use of bone conduction in navigation assistants has been explored in many studies recently (Kuriakose et al., 2023; Kuribayashi et al., 2022; Asakura, 2021). It would be interesting to examine how users feel while using it compared to standard headphones.

A guide dog has an advantage when it comes to helping the user to turn. In this study, we used analog clock-based navigation based on inspirations from Ross and Blasch (2000), Loomis et al. (2002), and instructions for turning right and left at 90 degrees. Despite successful pilot testing on this account, several participants described these methods as not optimal. Furthermore, it was also observed that most errors occurring during the experimental trials were related to incorrect turning. Hence we support the findings of Nakajima and Haruyama (2013), Miao et al. (2011) that clock-based directions could be challenging for people with visual impairments in navigation. However, it could be interesting to test with different methods for turning and directions. In addition, further studies should investigate the possibilities surrounding correction by the navigation solution itself while the user is navigating.

Another thing that could be relevant for future studies is a detailed exploration of navigation instructions based on vibrations. Several participants responded positively to using only vibration for navigation, as it would allow the opportunity to gather information from environmental sounds more significantly than if audio modality was used. It could be interesting to test the use of vibration as a unimodal navigation instruction or where vibration could be the predominant instruction method. In such scenarios, audio could supplement the information that cannot be conveyed through vibrations.

As mentioned in the limitations, this study involved only seven participants with visual impairments. The rest of the participants were blindfolded sighted participants because of the restrictions in place at the time of the experiment. The study can be further extended using the same experiment framework but with more participants with visual impairments. It could provide further insights into the problem domain.

Finally, the study suggests future development options for customization. The participants expressed a need to decide when they receive information, how they receive it, and how much they receive. Besides, few participants pointed out choosing the vibration pattern, as it would be easy to distinguish vibration from other applications such as incoming calls or messages. Therefore, customization could be interesting in future research. On the other hand, though customization is desirable, it is crucial that the solution still follows a standardized way of functioning. If it becomes too complex, the possibility of using the same will be reduced (Fosse et al., 2016).

7. Conclusion

This study aimed to determine how people with visual impairments performed in terms of time and errors encountered while navigating various types and modes of instruction, such as short and descriptive audio instructions and audio + vibration instructions. As a result of challenges in recruiting people with visual impairments due to social conditions, blindfolded sighted people were also the participants in the study. No significant differences were found in travel time or errors among participants who received short or detailed directions during navigation. The study also concluded that participants had no significant effect on their navigation performance even if they received multimodal instructions (audio + vibration) rather than audio-only instructions. While navigating, participants may or may not wish to hear details about the environment and obstacles, as the preference was found to be subjective. Participants with visual impairments spent less time and made fewer errors while navigating than blindfolded participants with 20/20 vision. This study provides insights into the navigation preferences of people with visual impairments. The study also suggests that a navigation assistant could be made more usable by customizing the type and mode of instructions for the users. The experiment framework described in this paper can be used in future studies with more people with visual impairments. We believe the findings from this study will be helpful in further developing effective navigation assistant systems.

CRediT authorship contribution statement

Bineeth Kuriakose: Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Ida Marie Ness:** Methodology, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft. **Maja Åskov Tengstedt:** Methodology, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft. **Jannicke Merete Svendsen:** Methodology, Validation, Investigation, Data curation, Visualization, Writing – original draft. **Terese Bjørseth:** Methodology, Validation, Investigation, Data curation, Visualization, Writing – original draft. **Bijay Lal Pradhan:** Validation, Formal analysis. **Raju Shrestha:** Conceptualization, Supervision, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to express their sincere gratitude to all the participants who generously contributed their time and valuable experiences to this study.

References

Ahmetovic, D., Guerreiro, J., Ohn-Bar, E., Kitani, K.M., Asakawa, C., 2019. Impact of expertise on interaction preferences for navigation assistance of visually impaired individuals. In: Proceedings of the 16th International Web for All Conference. pp. 1–9.

Amemiya, T., Sugiyama, H., 2010. Orienting kinesthetically: A haptic handheld wayfinder for people with visual impairments. ACM Trans. Access. Comput. (TACCESS) 3 (2), 1–23.

Anagnostakis, G., Antoniou, M., Kardamitsi, E., Sachinidis, T., Koutsabasis, P., Stavrakis, M., Vosinakis, S., Zissis, D., 2016. Accessible museum collections for the visually impaired: combining tactile exploration, audio descriptions and mobile gestures. In: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct. pp. 1021–1025.

Apostolopoulos, I., Fallah, N., Folmer, E., Bekris, K.E., 2014. Integrated online localization and navigation for people with visual impairments using smart phones. ACM Trans. Interact. Intell. Syst. (TiiS) 3 (4), 1–28.

Arditi, A., Tian, Y., 2013. User interface preferences in the design of a camera-based navigation and wayfinding aid. J. Vis. Impair. Blind. 107 (2), 118–129.

Asakura, T., 2021. Bone conduction auditory navigation device for blind people. Appl. Sci. 11 (8), 3356.

Bainbridge, W.S., 2004. Berkshire Encyclopedia of Human-Computer Interaction. Vol. 1. Berkshire Publishing Group LLC.

Barontini, F., Catalano, M.G., Pallottino, L., Leporini, B., Bianchi, M., 2020. Integrating wearable haptics and obstacle avoidance for the visually impaired in indoor navigation: A user-centered approach. IEEE Trans. Haptics 14 (1), 109–122.

Bella, M., Hanington, B., 2012. Universal Methods of Design. Rockport Publishers, Beverly, MA, p. 204.

Bouzit, M., Chaibi, A., De Laurentis, K., Mavroidis, C., 2004. Tactile feedback navigation handle for the visually impaired. In: ASME International Mechanical Engineering Congress and Exposition. Vol. 47063. pp. 1171–1177.

Brock, A.M., Truillet, P., Oriola, B., Picard, D., Jouffrais, C., 2015. Interactivity improves usability of geographic maps for visually impaired people. Hum.–Comput. Interact. 30 (2), 156–194.

Budrionis, A., Plikynas, D., Daniušis, P., Indrulionis, A., 2022. Smartphone-based computer vision travelling aids for blind and visually impaired individuals: A systematic review. Assist. Technol. 34 (2), 178–194.

Cosgun, A., Sisbot, E.A., Christensen, H.I., 2014. Evaluation of rotational and directional vibration patterns on a tactile belt for guiding visually impaired people. In: 2014 IEEE Haptics Symposium. HAPTICS, IEEE, pp. 367–370.

Dim, N.K., Ren, X., 2017. Investigation of suitable body parts for wearable vibration feedback in walking navigation. Int. J. Hum.-Comput. Stud. 97, 34–44.
Firmino, E., Teófilo, M., 2013. Visually impaired navigation assistant for emerging market using tactile floor, feature phone and audio descriptions. In: Proceedings of the 4th Annual Symposium on Computing for Development. pp. 1–2.
Fosse, P., Tellevik, J.M., Storliløkken, M., Martinsen, H., Grindheim, B.S., 2016. Utfordringer Knyttet Til Mobilitet for Personer Med Moderat Eller Alvorlig Synssvekkelse. En Teoretisk Og Praktisk Tilnærming Til Fagfeltet. Statped.
Gaver, W.W., 1991. Sound support for collaboration. In: Proceedings of the Second European Conference on Computer-Supported Cooperative Work ECSCW'91. Springer, pp. 293–308.
Golledge, R.G., Klatzky, R.L., Loomis, J.M., 1996. Cognitive mapping and wayfinding by adults without vision. In: The Construction of Cognitive Maps. Springer, pp. 215–246.
Golledge, R.G., Marston, J.R., Loomis, J.M., Klatzky, R.L., 2004. Stated preferences for components of a personal guidance system for nonvisual navigation. J. Vis. Impair. Blind. 98 (3), 135–147.
Grifoni, P., 2009. Multimodal Human Computer Interaction and Pervasive Services. IGI Global.
Groves, R.M., Fowler Jr., F.J., Couper, M.P., Lepkowski, J.M., Singer, E., Tourangeau, R., 2011. Survey Methodology. John Wiley & Sons.
Holland, S., Morse, D.R., Gedenryd, H., 2002. AudioGPS: Spatial audio navigation with a minimal attention interface. Pers. Ubiquitous Comput. 6 (4), 253–259.
Jacobson, R.D., 1998. Cognitive mapping without sight: Four preliminary studies of spatial learning. J. Environ. Psychol. 18 (3), 289–305.
Jadhav, A., Sonar, R., 2009. Analytic hierarchy process (AHP), weighted scoring method (WSM), and hybrid knowledge based system (HKBS) for software selection: a comparative study. In: 2009 Second International Conference on Emerging Trends in Engineering & Technology. IEEE, pp. 991–997.
Jaimes, A., Sebe, N., 2007. Multimodal human-computer interaction: A survey. Comput. Vis. Image Underst. 108 (1–2), 116–134.
Johannessen, A., Tufte, P.A., Christoffersen, L., 2010. Introduksjon Til Samfunnsvitenskapelig Metode. Vol. 4. Abstrakt Oslo.
José, J., Farrajota, M., Rodrigues, J.M.F., du Buf, J.M.H., 2011. The SmartVision local navigation aid for blind and visually impaired persons. Int. J. Digit. Content Technol. Appl. 5, 362–375.
Katz, B.F., Picinali, L., 2011. Spatial audio applied to research with the blind. Adv. Sound Localization 225–250.
Khan, S., Nazir, S., Khan, H.U., 2021. Analysis of navigation assistants for blind and visually impaired people: A systematic review. IEEE Access 9, 26712–26734.
Kolarik, A.J., Cirstea, S., Pardhan, S., Moore, B.C., 2014. A summary of research investigating echolocation abilities of blind and sighted humans. Hear. Res. 310, 60–68.
Kuriakose, B., Shrestha, R., Sandnes, F.E., 2020. Multimodal navigation systems for users with visual impairments—A review and analysis. Multimodal Technol. Interact. 4 (4), 73.
Kuriakose, B., Shrestha, R., Sandnes, F.E., 2022. Tools and technologies for blind and visually impaired navigation support: a review. IETE Tech. Rev. 39 (1), 3–18.
Kuriakose, B., Shrestha, R., Sandnes, F.E., 2023. DeepNAVI: A deep learning based smartphone navigation assistant for people with visual impairments. Expert Syst. Appl. 212, 118720.
Kuribayashi, M., Kayukawa, S., Vongkulbhisal, J., Asakawa, C., Sato, D., Takagi, H., Morishima, S., 2022. Corridor-Walker: Mobile indoor walking assistance for blind people to avoid obstacles and recognize intersections. Proc. ACM Hum. -Comput. Interact. 6 (MHCI), 1–22.
Lazar, J., Feng, J.H., Hochheiser, H., 2017. Research Methods in Human-Computer Interaction. Morgan Kaufmann.
Lee, Y.H., Medioni, G., 2016. RGB-D camera based wearable navigation system for the visually impaired. Comput. Vis. Image Underst. 149, 3–20.
Lewis, L., Sharples, S., Chandler, E., Worsfold, J., 2015. Hearing the way: requirements and preferences for technology-supported navigation aids. Applied Ergon. 48, 56–69.
Li, B., Munoz, J.P., Rong, X., Chen, Q., Xiao, J., Tian, Y., Arditi, A., Yousuf, M., 2018. Vision-based mobile indoor assistive navigation aid for blind people. IEEE Trans. Mob. Comput. 18 (3), 702–714.
Liu, G., Yu, T., Yu, C., Xu, H., Xu, S., Yang, C., Wang, F., Mi, H., Shi, Y., 2021. Tactile compass: Enabling visually impaired people to follow a path with continuous directional feedback. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. pp. 1–13.
Loomis, J.M., Lippa, Y., Klatzky, R.L., Golledge, R.G., 2002. Spatial updating of locations specified by 3-d sound and spatial language. J. Exp. Psychol: Learn. Mem. Cogn. 28 (2), 335.
Loomis, J.M., Marston, J.R., Golledge, R.G., Klatzky, R.L., 2005. Personal guidance system for people with visual impairment: A comparison of spatial displays for route guidance. J. Vis. Impair. Blind. 99 (4), 219–232.
Magnusson, C., Molina, M., Rassmus-Gröhn, K., Szymczak, D., 2010. Pointing for non-visual orientation and navigation. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries. pp. 735–738.
Manduchi, R., Coughlan, J., 2012. (Computer) vision without sight. Commun. ACM 55 (1), 96–104.
McDonald, A.M., Knight, R.C., Campbell, M.K., Entwistle, V.A., Grant, A.M., Cook, J.A., Elbourne, D.R., Francis, D., Garcia, J., Roberts, I., et al., 2006. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. Trials 7 (1), 1–8.
Miao, M., Spindler, M., Weber, G., 2011. Requirements of indoor navigation system from blind users. In: Symposium of the Austrian HCI and Usability Engineering Group. Springer, pp. 673–679.
Nakajima, M., Haruyama, S., 2013. New indoor navigation system for visually impaired people using visible light communication. EURASIP J. Wireless Commun. Networking 2013 (1), 1–10.
Nguyen, T.A., Kharitonov, E., Copet, J., Adi, Y., Hsu, W.-N., Elkahky, A., Tomasello, P., Algayres, R., Sagot, B., Mohamed, A., et al., 2023. Generative spoken dialogue language modeling. Trans. Assoc. Comput. Linguist. 11, 250–266.
Pasqualotto, A., Proulx, M.J., 2012. The role of visual experience for the neural basis of spatial cognition. Neurosci. Biobehav. Rev. 36 (4), 1179–1187.
Patel, R., 1995. Forskningsmetodikkens Grunnlag: Å Planlegge, Gjennomføre Og Rapportere En Undersøkelse. Universitetsforl.
Pielot, M., Boll, S., 2010. Tactile Wayfinder: comparison of tactile waypoint navigation with commercial pedestrian navigation systems. In: International Conference on Pervasive Computing. Springer, pp. 76–93.
Pielot, M., Poppinga, B., Boll, S., 2010. PocketNavigator: vibro-tactile waypoint navigation for everyday mobile devices. In: Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services. pp. 423–426.
Plikynas, D., Žvironas, A., Budrionis, A., Gudauskis, M., 2020. Indoor navigation systems for visually impaired persons: Mapping the features of existing technologies to user needs. Sensors 20 (3), 636.
Ponchillia, P.E., Jo, S.-J., Casey, K., Harding, S., 2020. Developing an indoor navigation application: Identifying the needs and preferences of users who are visually impaired. J. Vis. Impair. Blind. 114 (5), 344–355.
Quinones, P.-A., Greene, T., Yang, R., Newman, M., 2011. Supporting visually impaired navigation: a needs-finding study. In: CHI'11 Extended Abstracts on Human Factors in Computing Systems. ACM, pp. 1645–1650.
Real, S., Araujo, A., 2019. Navigation systems for the blind and visually impaired: Past work, challenges, and open problems. Sensors 19 (15), 3404.
Ross, D.A., Blasch, B.B., 2000. Wearable interfaces for orientation and wayfinding. In: Proceedings of the Fourth International ACM Conference on Assistive Technologies. pp. 193–200.
Sánchez, J., de la Torre, N., 2010. Autonomous navigation through the city for the blind. In: Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility. pp. 195–202.
Shoval, S., Borenstein, J., Koren, Y., 1994. Mobile robot obstacle avoidance in a computerized travel aid for the blind. In: Proceedings of the 1994 IEEE International Conference on Robotics and Automation. IEEE, pp. 2023–2028.
Spagnol, S., Wersényi, G., Bujacz, M., Bălan, O., Herrera Martínez, M., Moldoveanu, A., Unnthorsson, R., 2018. Current use and future perspectives of spatial audio technologies in electronic travel aids. Wirel. Commun. Mob. Comput. 2018.
Strothotte, T., 1995. MoBIC : user needs and preliminary design for a mobility aid for blind and elderly travelers, The European Context for Assistive Technology. In: The European Context for Assistive Technology: Proceedings of the 2nd TIDE Congress. Vol. 1. p. 348.
Tan, C.-H., Silva, A., Lee, R., Wang, K., Nah, F.F.-H., 2016. HCI testing in laboratory or field settings. In: International Conference on HCI in Business, Government, and Organizations. Springer, pp. 110–116.
Thaler, L., Goodale, M.A., 2016. Echolocation in humans: an overview. Wiley Interdiscip. Rev. Cogn. Sci. 7 (6), 382–393.
Tinti, C., Adenzato, M., Tamietto, M., Cornoldi, C., 2006. Visual experience is not necessary for efficient survey spatial cognition: evidence from blindness. Q. J. Exp. Psychol. 59 (7), 1306–1328.
Triantafyllopoulos, A., Schuller, B.W., İymen, G., Sezgin, M., He, X., Yang, Z., Tzirakis, P., Liu, S., Mertes, S., André, E., et al., 2023. An overview of affective speech synthesis and conversion in the deep learning era. Proc. IEEE.
Tsukada, K., Yasumura, M., 2004. Activebelt: Belt-type wearable tactile display for directional navigation. In: International Conference on Ubiquitous Computing. Springer, pp. 384–399.
Wang, H.-C., Katzschmann, R.K., Teng, S., Araki, B., Giarré, L., Rus, D., 2017. Enabling independent navigation for visually impaired people through a wearable vision-based feedback system. In: 2017 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 6533–6540.
Wilson, J., Walker, B.N., Lindsay, J., Cambias, C., Dellaert, F., 2007. Swan: System for wearable audio navigation. In: 2007 11th IEEE International Symposium on Wearable Computers. IEEE, pp. 91–98.
Xiao, J., Joseph, S.L., Zhang, X., Li, B., Li, X., Zhang, J., 2015. An assistive navigation framework for the visually impaired. IEEE Trans. HapticsMS 45 (5), 635–640.
Xu, S., Yang, C., Ge, W., Yu, C., Shi, Y., 2020. Virtual Paving: Rendering a smooth path for people with visual impairment through vibrotactile and audio feedback. Proc. ACM Interact. Mob. Wear. Ubiquitous Technol. 4 (3), 1–25.
Yatani, K., Banovic, N., Truong, K., 2012. SpaceSense: representing geographical information to visually impaired people using spatial tactile feedback. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 415–424.