MR-Sense: A Mixed Reality Environment Search Assistant for Blind and Visually Impaired People

1st Hannah Schieber
Friedrich-Alexander-Universität Erlangen-Nürnberg
Erlangen, Germany
Technical University of Munich School of Medicine and Health Klinikum rechts der Isar
Munich, Germany
hannah.schieber@fau.de

2nd Constantin Kleinbeck
Technical University of Munich School of Medicine and Health Klinikum rechts der Isar
Munich, Germany
Friedrich-Alexander-Universität Erlangen-Nürnberg
Erlangen, Germany
constantin.kleinbeck@tum.de

3rd Luisa Theelke
Technical University of Munich School of Medicine and Health Klinikum rechts der Isar
Munich, Germany
Friedrich-Alexander-Universität Erlangen-Nürnberg
Erlangen, Germany
luisa.theelke@fau.de

4th Miriam Kraft
Friedrich-Alexander-Universität Erlangen-Nürnberg Erlangen, Germany

5th Julian Kreimeier
Technical University of Munich School of Medicine and Health Klinikum rechts der Isar
Munich, Germany
julian.kreimeier@tum.de

6th Daniel Roth
Technical University of Munich School of Medicine and Health Klinikum rechts der Isar
Munich, Germany
daniel.roth@tum.de

***Abstract*—Search tasks can be challenging for blind or visually impaired people. To determine an object's location and to navigate there, they often rely on the limited sensory capabilities of a white cane, search haptically, or ask for help. We introduce MR-Sense, a mixed reality assistant to support search and navigation tasks. The system is designed in a participatory fashion and utilizes sensory data of a standalone mixed reality head-mounted display to perform deep learning-driven object recognition and environment mapping. The user is supported in object search tasks via spatially mapped audio and vibrotactile feedback. We conducted a preliminary user study including ten blind or visually impaired participants and a final user evaluation with thirteen blind or visually impaired participants. The final study reveals that MR-Sense alone cannot replace the cane but provides a valuable addition in terms of usability and task load. We further propose a standardized evaluation setup for replicable studies and highlight relevant potentials and challenges fostering future work towards employing technology in accessibility.**

***Index Terms*—blind and visually impaired people, accessibility, mixed reality**

Fig. 1: MR-Sense in a use case scenario. The head-mounted display allows for 3D sensing, deep learning-based object recognition, and object mapping. An obstacle detection warning provides distance-based vibrotactile feedback via a smartwatch for collision prevention. Audio feedback guides the user in search tasks.

I. Introduction

According to the World Health Organization (WHO), 295 million persons are impacted by significant visual impairment and 43 million persons are blind [1]. Visual impairment and blindness can have many characteristics, for example, a remaining light-dark distinction, or a restriction of color perception. The type of limitation and the time of the onset of the impairment is crucial. People with early blindness often find it easier to cope with everyday life challenges than people who previously relied on their sense of vision. These effects are becoming increasingly important because of the aging population [2]. Two common examples where the consequences of visual limitations become apparent are the search for objects [3], [4], and the collision-free navigation to them. Identifying objects outside the sensing area of their usual assistive aid [5] can become a major challenge for blind or visually impaired people (BVIP), as they need to haptically scan the environment or ask for help. Accidents often occur on the person's upper body, as this area is not sensed by common assistive aids [6]. Searching in an environment where the location of an object is unknown or forgotten is difficult, time-consuming, and often frustrating [3]. Assistive aids like the white cane offer immediate, analog audio-haptic feedback to assist BVIP for navigation. However, to locate and find an object additional haptic interaction via hands is required. In contrast, visual assistance services, to locate objects faster such as BeMyEyes [7], typically depend on external help and have limited availability.

Object search consists of two steps: first, locating the object, and second, safely navigating to it while avoiding obstacles. Existing approaches address mainly either the object localization aspect [8]–[10] or the navigation part [11]–[15] separately or identify one single object of interest in the current field of view (FoV), and navigate to it [16], [17].

A. Contribution

In this paper, we propose MR-Sense, a mixed reality (MR) audio and vibrotactile environment search assistant, developed in a participatory fashion with BVIP and evaluated in our standardized setup. We contribute a standardized setup to improve comparability and reproducibility among future studies for assistive systems.

MR-Sense utilizes a 3D-spatial scan of the environment to memorize located objects and provides non-contact obstacle warnings through gesture control and vibrotactile feedback. Our approach utilizes deep learning-based object recognition and 3D sensing, enabling it to identify and announce objects even beyond the current FoV of the head-mounted display (HMD). The system provides audio feedback with clockwise indication, distance, and height from the user's standpoint. Unlike smartphone-based approaches, the user's hands remain free when using MR-Sense, see Fig. 1. For the user, MR-Sense provides assistance during the hands-on search process by precisely identifying the location of objects. Thus, it minimizes the need for extensive haptic interaction with the environment, allowing the user to focus on sensing one specific object.

II. Related Work

Extending the sensing capabilities of assistive aids for BVIP has been addressed commercially [7], [18], [19] and in many research directions ([12]–[14], [20]–[22]). BVIP most often use their white cane or a guide dog for navigation [5]. To extend their reality using a HMD, they would prefer an adjustable, context-sensitive and customized approach [23].

A. Obstacle Avoidance and Navigation

While extending the white cane [12], [13], [22] is most reasonable, visual processing enabled by cameras can assist the users too. Zoller et al. [14], [20] use an RGB-D camera (Microsoft Kinect) and provide audio and haptic feedback in the case of obstacles. Chaccour and Badr [24] proposed an approach, where a smartphone interacts with a remote server, markers and indoor cameras to help BVIP navigate. Hsieh et al. [25] utilize a glove to provide vibrotactile feedback for navigation. In another work, Hsieh et al. [26] combine audio feedback, a smartphone and a haptic glove for outdoor navigation.

To tackle the challenge of stair navigation, Zhao et al. [27] utilized projection-based augmented reality (AR), smart-glasses, and sonification. While previous approaches were, for example, marker-based [20], [24], nowadays deep learning-based approaches are used more frequently [15], [28]. Hsieh et al. [28] utilized an RGB-D camera and a convolutional neural network (CNN) to perform semantic segmentation and find a walkable path, such as outdoor sidewalks and crosswalks. Zhang et al. [15] used images captured by a HMD and fed them into a transformer-based approach, executed on a portable GPU, to identify a walkable path and transparent parts in the environment like glass doors. Katz et al. [16] addressed finding, for example, traffic lights using geolocalization and object identification.

B. Object Identification and Localization

Navigation and obstacle avoidance is one part of safely navigating to a desired object, the other part is identifying the object of interest. Jafri et al. [29] addressed object search utilizing camera-based object detection and RFID tags. Gautam et al. [30] identified objects in images and provided proximity navigation towards them via audio. Eckert et. al. [9] and Schieber et al. [10] used a keyword to scan the whole image currently captured by a HMD to identify objects. Lee et al. [17] optimized object localization using hand cues to identify objects near the user's hand. Chen et al. [31] combine hand tracking and object recognition to announce distances indicating arm length. Huppert et al. [8] proposed a different method for object grasping using a drone and haptic feedback in unknown environments. Instead of deep learning-based object identification, they utilized an OptiTrack system to identify objects.

III. Preliminary Interviews

The focus of the initial qualitative interviews was to i) identify what kind of assistive aids BVIP use, ii) how they handle an object search, iii) if they would be interested in using a MR-based audio and vibrotactile guidance assistant, and iv) how they would like to interact with such a system. For the latter, a preliminary prototype of MR-Sense and potential features were orally described. Each interview took 30 to 45 minutes on average.

A. Participants

To better understand BVIP's daily challenges and allow for a design process, we conducted semi-structured qualitative phone interviews with $N = 10$ participants ($M = 57.7$, $SD = 14.99$ years, five male, five female). Seven participants were blind and three were visually impaired. Two of the blind participants were sensitive to light.

B. Preliminary Results

Regarding our focus on object search (e.g., moved objects by others or forgetting where the object is), the participants either ask others for help ($N = 2$) or search for them ($N = 5$). In familiar environments such as one's own home, objects can be inadvertently misplaced or rearranged by guests ($N = 3$). The object search was pointed out as time-consuming and cumbersome. ("I spend a lot of time searching for things." and "Searching is sometimes frustrating."). For navigation, the majority appreciates the white cane ($N = 8$).

After explaining our idea we asked whether they would like to have such an assistive system and how they would like to interact with it and receive feedback. Sensing objects outside of the sensing area of a white cane was mentioned as useful. In general, they pointed out that such a system can improve independence as they are often irritated by people yelling "Attention, attention!" ($N = 3$), as they don't know what to be attentive to. To activate the system, voice activation ($N = 6$) or multiple activation options ($N = 3$) could be useful. They further mentioned that an activation via touch is difficult, as one hand is already busy with the white cane. In terms of making distances audible, distances larger than ten meters were classified as hard to imagine ($N = 2$) while distances below ten meters were noted as sensible to imagine ($N = 7$). The combination with a direction indication was stated as suitable and specifically the indications by means of a clock were suggested ($N = 3$).

C. Derived Design Choices

A system like MR-Sense could be useful in daily life as object search was noted as time-intensive and cumbersome *(D1)*. Since it was pointed out that voice activation could be beneficial, we determined to use this as the initial activation for the system *(D2)*. We decided to identify objects and map them on a 3D map, so that object search can be supported even if the object is not in the current FoV of the HMD *(D1)*. To announce an object's position clockwise indications and distances in meters were noted as good to imagine *(D3)*. Furthermore, it was mentioned that objects, that are outside the sensing range of the white cane, are a potential risk *(D4)*. For this reason, we add vibrotactile feedback via a smartwatch as a distance indicator to support the user earlier with warnings about potential obstacles *(D4)*. Based on our interviews, we concluded that:

- *(D1)* A system to support object search in known and unknown environments is useful ($N = 7$)
- *(D2)* The use of voice activation would be beneficial ($N = 6$)
- *(D3)* Distance indications using clockwise indications are useful ($N = 3$)
- *(D4)* The system should be aware of other obstacles while navigating to the desired object ($N = 3$)

D. Hypotheses

Our system, MR-Sense supports object search and navigation. Searching for objects can be frustrating especially when the white cane cannot sense the object. Our system can identify objects if they were captured at some point by the HMD camera, thus we hypothesize **H1**: *Our system will reduce the search time and walked distance in object search tasks.*

Moreover, MR-Sense will only be used for a short time period and introduces additional feedback compared to the white cane. We hypothesize **H2**: *The use of our (unfamiliar) system will lead to a higher workload compared to the white cane as daily assistive aid.*

Based on the information gained from the preliminary discussion, the white cane is very useful and valuable for BVIP. Thus, we hypothesize using the combination of MR-Sense and the white cane will result in a higher usability and user experience than purely using our system, which leads to **H3**: *The combination of the white cane and MR-Sense will score higher compared to only using our system in terms of usability and user experience.*

IV. System Implementation

MR-Sense is built using *Unity*, *Python* and *Swift* and runs on a HMD (*HoloLens 2*), a smartwatch (*Apple Watch Series 6*), and an external server[1]. Fig. 2 depicts MR-Sense.

A. Object Localization

User interaction with the system primarily relies on voice input *(D2)*. Voice commands initiate object searches using class names as keywords, like *laptop* or *umbrella*. Upon detecting the object's position within the HMD's sensory data, the system provides audio feedback in a clockwise indication, along with distance and height information for the searched object *(D3)*, see Fig. 2.

Internally, MR-Sensecompares the stated class name with recognized objects. If the object hasn't been recognized yet, the class name is saved, and the system announces, "Your object was not found yet. As soon as it is noticed, you will be informed." Once the object is identified, its location is announced. If multiple objects of the same class are detected, the latest one is reported. We employ the Yolo5s [32], a real-time processing model, for object detection using pretrained weights from the COCO dataset [33]. We set a threshold for YOLO5s to 35%/45% (confidence/IoU). We tested with different person heights and head orientation (from 1.50 meters up to 1.80 meters). The confidence ranged depending on the distance of user to the object from 76% at three meters up to 91% at one meter. For distances over three meters, the confidence decreased.

In detail, an external server hosts a RESTful API for querying information. When the application starts, the HMD sends images to the server. The server runs Yolo5s, returning the detected objects' class names, x- and y-center coordinates to the HMD. These 2D object center coordinates are projected back into the HMD camera coordinate system. Our system utilizes the HMD's 3D map and combines it with object detection results using a *sphere cast*. This process approximates the 3D position of the object. Based on the confidence of the object localization, spatial mesh limits were set to three meters in x, y, and z axes. However, object positions beyond this threshold can still be identified if the object was previously mapped.

[1]The source code will be published with the paper

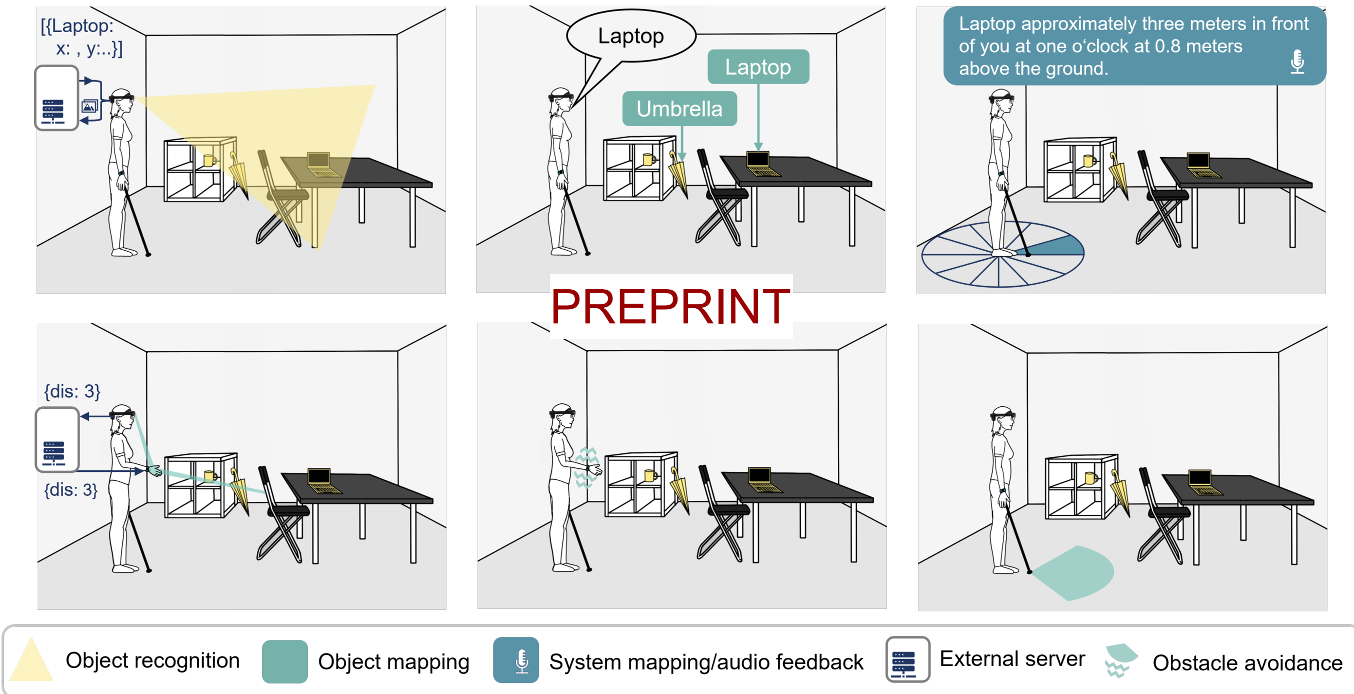

Fig. 2: **Overview of MR-Sense**. A server handles incoming requests with images and distance measures (top left, bottom left). The object which is searched can be announced via its class name (top center). MR-Sense announces its approximated location (top right). With the distance measure, the person can get an understanding of distances on their wrist via vibrotactile feedback on the smartwatch (bottom center) or additionally use the white cane to detect obstacles on the ground (bottom right).

The system latency was measured at 400ms: 200ms from the server and object detection model, with the remaining time used by Wi-Fi and client communication.

B. Obstacle Avoidance

Apart from the voice and audio interaction, the system provides vibrotactile feedback via the smartwatch on the user's hand, see Fig. 2 *(D4)*. We utilize the built-in option of the HMD to obtain a 3D mesh of the surroundings. The HMD's hand tracking feature enables detecting objects in front of the hand. When extending the hand forward, raycasts from the hand to the computed spatial mesh are used to determine the distance to the closest object (e.g., a wall). The distance from the collision is then handed to the external server and pushed to the smartwatch on the user's hand. Distances are indicated by different levels of vibration intervals to provide the user with an impression of distance. The vibration intervals are grouped in four different strengths: very close objects (all 400 ms, distances $< 0.5m$), close objects (all 800 ms, distances $< 0.5m - 0.8m$), normal (all 1200 ms, distances $< 0.8m - 1.5m$), and away (all 1600 ms, distances $< 1.5m - 2m$).

V. User Study

A. Design

To measure the usability and task load when locating objects and navigating to their place, we conducted a within-subject user study with BVIP. The participants were asked to complete a search task either in a baseline condition (Cane) with their white cane, with MR-Sense only (MixedReality), or with the combination of the white cane and MR-Sense (Combined). In the cane condition (Cane) they use their standard tool, the white cane for navigating. For the MR condition (MixedReality) they only use MR-Senseand in the Combined condition they use the white cane and MR-Sense.

B. Reproducible Study Apparatus

As a realistic and reproducible experimental setup, we prepared a common indoor room with two shelves, a table and a chair. All items were purchased at *IKEA*[2]. A banana, a laptop, a bottle and a flower were chosen as test objects to be located (these four objects are within the 80 categories of the COCO dataset mentioned above). As search objects, we selected a banana, a laptop, a bottle and a flower (these four objects are within the 80 categories of the COCO dataset mentioned above).

C. Study Procedure

Participants were welcomed and provided a detailed oral briefing about the study using a standardized form. After securing their consent, we collected demographic details, assessed their level of visual impairment, inquired about their experience with assistive aids, their familiarity and commitment to virtual, augmented, or MR technologies [34], see Fig. 4.

[2]KALLAX, white, 77x39x77 cm; KALLAX, white, 42x39x112 cm; LINNMON / ADILS Desk, darkgrey/black, 100x60x73 cm; GUNDE Chair, black

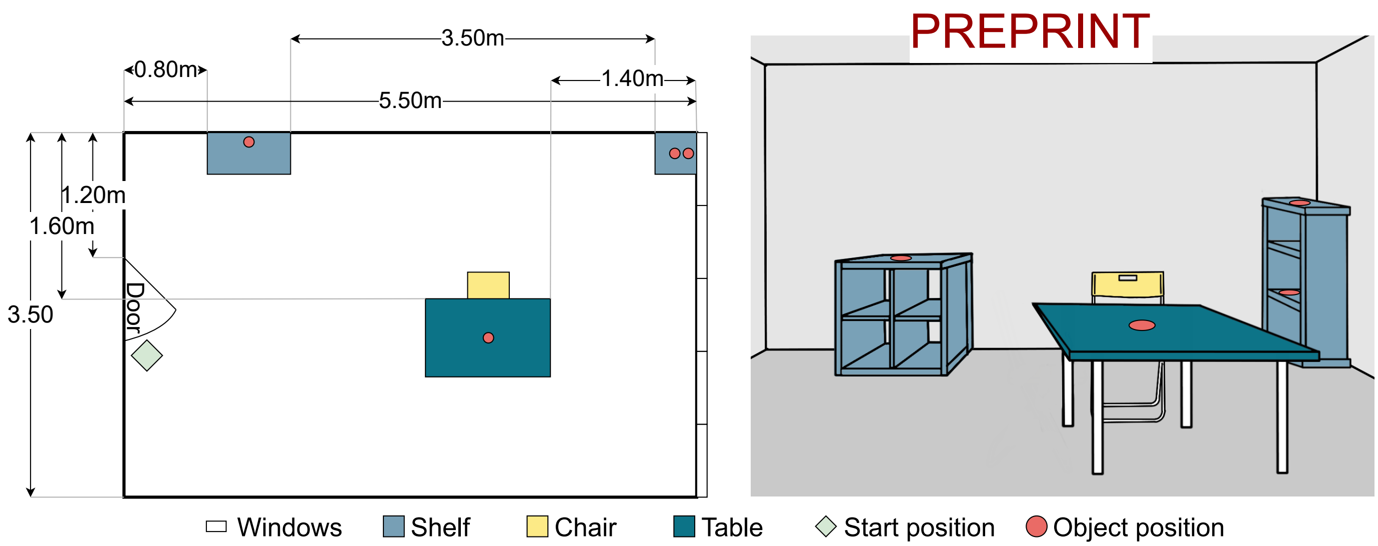

Fig. 3: Floor plan of the test room (left) and highlighted object positions in red (left and right). The room measures 5.50m $\times$ 3.50m and contains two shelves, a chair and a table. All items were purchased at *IKEA*.

Before the trials, participants explored the room for ten minutes using the white cane. They were then introduced to MR-Sense, featuring vibrotactile feedback and the HMD, followed by a brief familiarization phase. Upon clarifying any remaining doubts, experimental trials commenced with randomized conditions to prevent bias.

In each trial, participants were required to locate four objects within our standardized room, with the object orders randomized across conditions for each participant. When using MR-Sense, the application reset for each trial, allowing individual mapping without prior system preparation. The sequence of conditions (Cane, MixedReality, Combined) was randomly determined while ensuring a balanced allocation, though an additional occurrence of Combined, MixedReality, Cane was included due to an odd participant count. After participants familiarized themselves with the room, the four objects were placed in unknown locations, changing after each condition while participants were absent.

Upon successfully locating the objects in a condition, time taken and distance covered were recorded. Following this, the experimenter completed questionnaires, integrating qualitative feedback to provide context to the quantitative data. The experiment concluded upon completing all conditions.

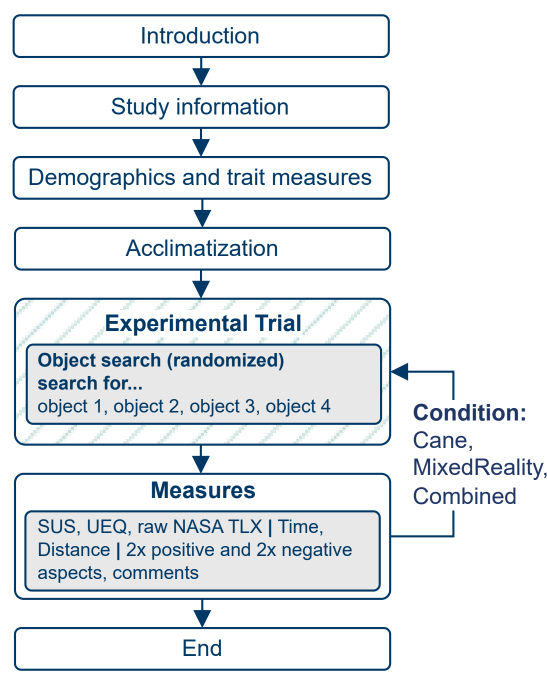

Fig. 4: Study Procedure with the individual phases introduction, study information, and questionnaires assessed at different points of the experiment.

D. Measures

1) Subjective Measures: To measure the system's usability and the participants' user experience and task load, we assessed the system usability scale (SUS) [35], the User Experience Questionnaire (UEQ) [36] and the raw NASA-TLX [37], [38] questionnaire. To analyze the context of these data with a covariate, the technology commitment was measured using the technology commitment questionnaire [34] consisting of twelve items each rated on a scale of one to five (ranging from strongly disagree to strongly agree). For the SUS we followed the standard procedure to rate the answers from one to five (strongly disagree = 1, strongly agree = 5). The UEQ was also answered based on the standard procedure provided in the UEQ handbook (-3 to +3). As the participants responded verbally, the raw NASA-TLX was answered in percentage points (0% up to 100%) instead of ticks at scales.

2) Objective Measures: Objective data was collected to evaluate performance by measuring the *Time* it took users to complete each search task and the *Distance* they walked to find all objects. This way, we were able to derive information

TABLE I: Participants' relevant data, i.e. their visual acuity (no remaining vision (NV), right eye (RE), left eye (LE), both eyes (BE), or FoV), used assistive aids, technology commitment (TC) and virtual (VR), augmented (AR) or MR (MR) experience denoted as XR. For the sake of anonymity and since all participants were from the same area, we excluded age and gender from the table.

Participant	Visual acuity	Color Detection	Braille-display	Reading aid	Smartphone	Speaking Devices	Navigation belt	White cane	Guide person	Guide dog	TC	XR experience
A	NV	-	-	✓	✓	✓	-	✓	-	-	35	-
B	NV	✓	✓	✓	-	✓	-	✓	-	-	36	-
C	NV	-	✓	-	✓	✓	-	✓	✓	-	37	-
D	NV	✓	-	✓	✓	-	-	✓	✓	-	39	✓
E	NV	✓	✓	✓	✓	✓	-	✓	✓	✓	29	✓
F	NV	✓	✓	✓	✓	✓	✓	✓	✓	-	42	✓
G	NV	✓	✓	✓	✓	✓	-	✓	✓	-	30	✓
H	NV	✓	✓	✓	✓	✓	-	✓	✓	✓	33	✓
I	RE 1%, LE 2%	-	✓	✓	✓	✓	-	✓	-	-	31	✓
J	BE $\leq$ 2%	-	✓	-	✓	✓	-	✓	✓	-	22	✓
K	RE $\leq$ 5%, LE $\leq$ 15%	-	-	-	✓	-	-	✓	-	-	35	✓
L	BE $\leq$ 2%	-	✓	✓	✓	✓	-	✓	✓	-	30	✓
M	FoV$\leq$ 5°	✓	✓	✓	✓	✓	✓	✓	✓	-	42	✓

about the temporal and spatial effectiveness of MR-Sense. The walking distance was measured based on the HMD position. Data acquisition began when a start signal was sent to the HMD, and upon completion of the participant's task, the recording was stopped by the experimenter's command.

3) Qualitative Comments: During and after the trials, the experimenter noted the participants' statements for each search task and confirmed these logs with the participant at the end of the condition. We further asked the participant to comment on two positive and two negative aspects of the respective condition (i.e., Cane, MixedReality, or Combined) and the experimenter noted these comments.

E. Participants

In total $N = 13$ BVIP between 25 and 68 years ($M = 50.08$, $SD = 14.99$ years, six male and seven female) were recruited via contacting associations for BVIP by posting the call for participants in their monthly newsletter. Potential participants called the authors and brief phone interviews were conducted. The participants were asked about their level of visual impairment and informed about the approximate time the experiment will take. The characteristics for their level of visual impairment varied, see Table I. Some participants were blind from birth ($N = 2$), others suffered from retinal detachment ($N = 3$), glaucoma ($N = 2$), cataract ($N = 3$), or a combination of multiple diseases ($N = 5$) leading to a low or zero level of vision. Participants who did not meet the requirements were not included in the evaluation. One of the participants with reaming vision ($N = 5$) counted as visually impaired while the others are legally blind. We conducted a *Landolt-C* vision acuity to ensure that all participants had an equivalent perception and do not have any individual advantages/disadvantages. Participants with light perception were able to localize the windows in the room because of the

TABLE II: Results of the individual questionnaires system usability scale, UEQ, and raw NASA-TLX. We report the median for each condition as well as $\chi^{2(2)}$and p

Subjective Measures	Mdn_{Cane}	$Mdn_{MixedReality}$	$Mdn_{Combined}$	$\chi^{2(2)}$	p
SUS	**85.00**	77.50	**85.00**	1.920	0.382
Raw NASA-TLX	18.33	**13.33**	20.00	0.695	0.706
Mental demand	30.77	25.77	**25.00**	0.050	0.706
Physical demand	26.15	17.69	**14.62**	1.117	0.572
Temporal Demand	14.62	13.85	**13.46**	0.066	0.967
Performance	**15.00**	18.84	17.69	2.137	0.343
Effort	29.62	**13.33**	19.23	1.555	0.459
Frustration	10.38	**9.23**	10.38	0.300	0.861
UEQ					
Attractiveness	1.67	2.00	**2.17**	2.130	0.344
Perspicuity	2.50	2.50	**2.75**	3.534	0.170
Efficiency	1.50	**2.00**	1.75	0.285	0.866
Dependability	1.50	1.75	**2.00**	0.136	0.934
Stimulation	1.75	2.00	**2.50**	4.044	0.132
Novelty	-1.00	2.00	**2.25**	22.167	$< \mathbf{0.001}$
Objective Measures					
Walking Time	**105.29**	201.62	162.49	6.615	**0.036**
Walking Distance	**27.39**	38.07	30.66	4.044	0.132

light influence. However, the objects for the search task were not perceptible for them at a distance of more than 30 cm.

In terms of everyday life aids, many participants use a color detection device ($N = 7$), braille display ($N = 10$), a navigation belt ($N = 2$), or the white cane ($N = 13$), see supplementary material. Regarding measuring their technology commitment, most of our participants had a moderate technology acceptance resulting in a mean of ($M = 33.83$, $SD = 5.29$, max $= 42$, max $= 22$). As listed in the supplementary material some participants already had experience with virtual, augmented or MR. Some participated in other experiments using virtual reality ($N = 6$), and others tested commercial glasses ($N = 3$) or assistive aids like a backpack stacked with multiple sensors ($N = 2$).

F. Analysis Approach

To analyze the differences between the three conditions, we performed Friedman tests and ANOVA. For pairwise comparisons, we used non-parametric Wilcoxon signed-rank or parametric t-tests, depending on the data distribution. If the Shapiro-Wilk test indicated a non-normal distribution we applied the Friedman Test, and if the result revealed a normal distribution we used ANOVA.

VI. Results

A. Subjective Measures

For the SUS, all conditions were rated above 68 which is categorized as above average [35]. The total scores varied between Cane and the other conditions, but according to the results, see Table II, the differences were not statistically significant ($p > 0.05$).

The total scores of the raw NASA-TLX showed that our defined search task and all three conditions have a medium level of workload (medium $10 - 29$). As reported in Table II, neither the total score of the raw NASA-TLX nor the individual scales reveal significant differences when assessing the Friedman test ($p > 0.05$).

In terms of *Novelty*, a statistically significant difference ($p < 0.001$) was found. Pairwise comparisons using Wilcoxon

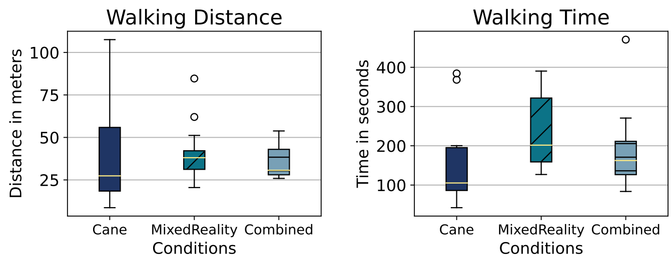

Fig. 5: Results of the conditions Cane, MixedReality, and Combined for walking distance and walking time.

signed-rank tests with a Bonferroni correction for multiple comparisons showed statistically significant differences in the *Novelty* score. Specifically, the *Novelty* score was significantly different between the conditions Cane/MixedReality ($Z = -3.149, p = 0.002$) and Cane/Combined ($Z = -3.150, p = 0.002$).

In addition to the Friedman tests, ANOVA was conducted for repeated measurements. These analyses confirmed the previously described findings. Furthermore, ANCOVA was used to analyze the impact of age, order of conditions, onset time of visual impairment or blindness, and technology commitment. However, no statistically significant results were found.

B. Objective Measures

Fig. 5 shows that the Combined (35.49 ± 9.99) condition resulted in a reduction in the mean and standard deviation of the walking distance compared to the Cane (37.51 ± 28.25) or MixedReality (40.61 ± 17.40) condition. The mean±standard deviation of time taken (in seconds) for Cane was 150.20 ± 110.51, for Mixed Reality it was 239.57 ± 98.08, and for the condition Combined it was 183.22 ± 102.90. A shorter completion time for the search task was observed in the condition Cane when compared to MixedReality or Combined conditions. However, a statistically significant difference was noted only between the Cane and MixedReality condition ($p = 0.036$). A Shapiro-Wilk test confirmed non-normal distribution, therefore, we assessed a Wilcoxon signed-rank test with a Bonferroni correction for multiple comparisons indicating a significant effect for the measured walking time for the condition Cane and MixedReality ($Z = -2.831$, $p = 0.005$). In addition to the Friedman tests, we analyzed the data using ANOVA for repeated measurements. The results of the analyses corroborated the above-stated findings.

C. Qualitative Comments

To ensure anonymity, the participants are named with randomized numbers. Most of our participants pointed out that the combination of our system and a white cane is helpful and adds an additional dimension to the search field (N=10). P5 mentioned that *“With the glasses and heights indication I don’t have to search everywhere from top to bottom”* and P12 told *“I get a new dimension for the search task”*. Participants with a remaining vision (P10 and P11) pointed out that a large directional indicator in the form of an arrow would be helpful to find the objects. P3 said *“I would like to get my walking direction painted towards the searched object”*. Participants P1, P7, P9, and P11 mentioned that the vibrotactile feedback could be stronger, for example, P7 said *“When I am concentrated, the vibration was too soft. I often did not notice it.”*. Additionally, P9 criticizes the latency as *”the vibration does not change fast enough when I move quickly”*. Other participants (e.g., P5, P6, P7, and P10) mentioned that the current HMD is too heavy but the system is overall useful.

The participants mentioned that with the white cane, they always have to memorize the whole room, while with our system their cognitive load felt more focused on listening to the clockwise and directional indications. P13 said *“When I walk only with the cane, I have to remember everything, with the glasses I only have to remember the clock”*. Moreover, they found the height information very useful, P8 said *“The height information helps with the search.”* and P13 agrees *“The information about the height is useful”*.

VII. Discussion

MR-Sense was developed in a participatory fashion and aims to support BVIP in indoor search tasks. In our user study, the participants performed a search task being exposed to a reproducible study apparatus that allows for the replication of the study and benchmarking of future systems. We measured the usability (SUS), user experience (UEQ), the workload (raw NASA-TLX), taken time, walked distance, and qualitative comments to observe the acceptance, advantages and disadvantages of MR-Sense.

Our system will reduce the search time and walked distance in object search tasks (**H1**). Regarding walking distance, we can confirm quantitatively that the distance was shorter and exhibited less variance in the Combined condition. This suggests that the directional indication aids targeted search and reduces the need for scanning the entire area while the white cane supports walking towards this direction. However, our results can not confirm this hypothesis statistically. We observed a significant difference in measured walking time between the baseline and the use of MR-Sense alone. Participants tended to wait for MR-Sense to complete the voice output, resulting in increased time compared to using the white cane alone. Therefore, faster and more precise audio feedback is crucial. Additionally, we observed that the Combined condition reduced walking time compared to MR-Sense alone.

We further hypothesized, *the use of our (unfamiliar) system will lead to a higher workload compared to the white cane as daily assistive aid* (**H2**). Regarding this hypothesis, the results do not show any significant differences. The median score is highest when using MR-Sense alone, while the Combined condition has lower ratings. However, MR-Sense alone shows less scatter compared to the baseline. Surprisingly, the Combined condition has the lowest task load, but the differences are not significant. Therefore, MR-Sense does not significantly increase the task load. One can conclude, that MR-Sense was received as a helpful add-on.

Our results do not statistically support that *the combination of the white cane and MR-Sense will score higher compared to only using our system in terms of usability and user experience* (**H3**). However, the results do not show any negative impacts either and indicate that the participants generally rate the MR-Sense system as well as the combination rather positively. In particular, the combination of MR-Sense and the white cane receives a comparable score in SUS compared to the baseline. One can argue that the familiar assistive aid supports the system usability as the participant feels safer than when only using a novel system. Yet, the novel system as a standalone also shows a SUS score above 68. The UEQ results for *Attractiveness*, *Perspicuity*, *Dependability*, *Stimulation* were rated higher by the participant for the condition Combined, but not at a significant level. In terms of *Attractiveness*, a higher scoring for Combined is observed, similar to using MR-Sense only (MixedReality).

Qualitative feedback during the experiments indicated that for some users, using only MR-Sense helps the perceived efficiency since focusing on only one system makes the task easier. For the condition Combined, the focus lies on two systems, while for the baseline, the participants constantly have to remember the whole room setup and possible search areas. Also, high scores are observed for *Stimulation* using both conditions containing MR-Sense. In the UEQ, *Stimulation* is assessed using items such as "motivating", "valuable", "interesting" or "exciting", and indicates that the participants were, on average, positive toward MR-Sense. The conditions with MR-Sense lead to a significant increase in the UEQ *Novelty* rating.

VIII. Limitations

Due to the small number of participants ($N = 13$), our results must be generalized with caution. It is worth noting that the prevalence of BVIP in the general population is only around 6.7% in Europe and 6.0% in the United States [1], [39]. Especially blind people are rare in society with below 0.1% in Germany and 0.2% in Russia [40]. Previous studies in this research area have also reported comparable sample sizes [8], [15], [41], [42]. Nonetheless, we emphasize the importance of investigating the needs of these smaller populations, facilitating their enhanced participation in a society primarily tailored to the sighted majority. Our study offers initial insights upon which future research can build to further enhance the system.

Also, many participants expressed great interest in MR-Sense. Learning and being able to use the system in medium to long term requires more extensive studies. Also, using the well-known white cane as a baseline is a challenging comparison as some participants have been using it since their teenage years.

In terms of object recognition, MR-Sense currently identifies objects on a category level, so the most recent one per category is mapped and mentioned.

IX. Future Work

Future user studies should potentially increase the sample and further investigate the user interactions, e.g., number of system activations, learnability, and (social) acceptability, as well as long-term effects. Considering a ground truth without an assistive aid could be of interest. However, we designed our system as an addition, rather than a replacement of tools such as the white cane. For future evaluations, our reproducible study apparatus could be used with other assistive approaches or extensions of MR-Sense addressing and comparing the accessibility for BVIP. Our user study setup is easily extendable, by adding more objects or additional furniture to test in a more cluttered setup. Regarding the technical implementation, the vibrotactile feedback and the HMD's bulky hardware should be optimized. Furthermore, visual feedback can be opted-in (e.g., an AR arrow pointing to the location or a flashing contour) considering the severity of visual impairment. Moreover, as head orientation can be measured, future adaptations of MR-Sense could notify the user if the head position is too adverse to recognize objects.

X. Conclusion

In this paper, we presented MR-Sensea consumer hardware-based MR environment sensing assistant. The system utilizes deep learning and 3D spatial understanding to provide multimodal feedback for BVIP in order to support them in search and navigation tasks. MR-Sense was developed in participatory fashion, ensuring that the BVIP perspectives and needs were considered. To promote reproducibility and comparability in future research, we also present a standardized study apparatus. We used this setup to evaluate the performance of our system in a practical search and navigation task. In our final study with BVIP, we found that overall, the subjective usability, user experience, and task load measures showed positive results. Moreover, we reached similar ratings and performance results to the white cane as the state-of-the-art baseline. Our approach showcases relevant potentials and challenges fostering following work towards employing technology in accessibility.

Compliance with Ethical Standards

The studies involving human participants were reviewed and approved by the Ethics Committee at Friedrich-Alexander-Universität Erlangen-Nürnberg and were conducted in accordance to the Declaration of Helsinki. All participants gave their consent for this study.

Acknowledgment

We thank all participants for their time and their helpful feedback when taking part in the user study.

References

[1] WHO, "Blindness and vision impairment," 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

[2] R. Varma, T. S. Vajaranant, B. Burkemper, S. Wu, M. Torres, C. Hsu, F. Choudhury, and R. McKean-Cowdin, "Visual impairment and blindness in adults in the united states: demographic and geographic variations from 2015 to 2050," *JAMA ophthalmology*, pp. 802–809, 2016.

[3] S. Lee, C. W. Yuan, B. V. Hanrahan, M. B. Rosson, and J. M. Carroll, "Reaching out: Investigating different modalities to help people with visual impairments acquire items," in *International ACM SIGACCESS Conference on Computers and Accessibility*. New York, NY, USA: Association for Computing Machinery, 2017, pp. 389–390.

[4] W. Jeamwatthanachai, M. Wald, and G. Wills, "Indoor navigation by blind people: Behaviors and challenges in unfamiliar spaces and buildings," *British Journal of Visual Impairment*, vol. 37, no. 2, pp. 140–153, 2019.

[5] M. A. Williams, A. Hurst, and S. K. Kane, ""Pray before you step out" describing personal and situational blind navigation behaviors," in *International ACM SIGACCESS Conference on Computers and Accessibility*. New York, NY, USA: Association for Computing Machinery, 2013, pp. 1–8.

[6] R. Manduchi and S. Kurniawan, "Mobility-related accidents experienced by people with visual impairment," *Insight: Research and Practice in Visual Impairment and Blindness*, no. 2, pp. 44–54, 2011.

[7] BeMyEyes. (2022) BeMyEyes. BeMyEyes. [Online]. Available: https://www.bemyeyes.com/

[8] F. Huppert, G. Hoelzl, and M. Kranz, "GuideCopter - a precise drone-based haptic guidance interface for blind or visually impaired people," in *Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, 2021.

[9] M. Eckert, M. Blex, C. M. Friedrich, and others, "Object detection featuring 3d audio localization for microsoft HoloLens," in *Int. Joint Conf. on Biomedical Engineering Systems and Technologies*, vol. 5, 2018, pp. 555–561.

[10] H. Schieber, C. Kleinbeck, C. Pradel, L. Theelke, and D. Roth, "A mixed reality guidance system for blind and visually impaired people," in *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 2022, pp. 726–727.

[11] S. Hashino and R. Ghurchian, "A blind guidance system for street crossings based on ultrasonic sensors," in *The 2010 IEEE International Conference on Information and Automation*. IEEE, 2010, pp. 476–481.

[12] B. Kuriakose, R. Shrestha, and F. E. Sandnes, "Tools and Technologies for Blind and Visually Impaired Navigation Support: A Review," *IETE Technical Review*, vol. 39, no. 1, pp. 3–18, 2020.

[13] A. D. P. dos Santos, F. O. Medola, M. J. Cinelli, A. R. Garcia Ramirez, and F. E. Sandnes, "Are electronic white canes better than traditional canes? A comparative study with blind and blindfolded participants," *Universal Access in the Information Society*, pp. 93–103, 2021.

[14] H. Liu, J. Wang, X. Wang, and Y. Qian, "Isee: Obstacle detection and feedback system for the blind," in *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, ser. UbiComp/ISWC'15 Adjunct. New York, NY, USA: Association for Computing Machinery, 2015, p. 197–200. [Online]. Available: https://doi.org/10.1145/2800835.2800917

[15] J. Zhang, K. Yang, A. Constantinescu, K. Peng, K. Müller, and R. Stiefelhagen, "Trans4Trans: Efficient Transformer for Transparent Object Segmentation To Help Visually Impaired People Navigate in the Real World," in *International Conference on Computer Vision (ICCV) Workshops*, 2021, pp. 1760–1770.

[16] B. F. G. Katz, S. Kammoun, G. Parseihian, O. Gutierrez, A. Brilhault, M. Auvray, P. Truillet, M. Denis, S. Thorpe, and C. Jouffrais, "NAVIG: augmented reality guidance system for the visually impaired," *Virtual Reality*, pp. 253–269, 2012.

[17] K. Lee and H. Kacorri, "Hands holding clues for object recognition in teachable machines," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–12.

[18] A. Cosgun, E. A. Sisbot, and H. I. Christensen, "Evaluation of rotational and directional vibration patterns on a tactile belt for guiding visually impaired people," in *2014 IEEE Haptics Symposium (HAPTICS)*. IEEE, 2014, pp. 367–370.

[19] I. Cloudsight. (2022) taptapseeapp. Cloudsight, Inc.

[20] M. Zöllner, S. Huber, H.-C. Jetter, and H. Reiterer, "NAVI – A Proof-of-Concept of a Mobile Navigational Aid for Visually Impaired Based on the Microsoft Kinect," in *Human-Computer Interaction – INTERACT 2011*. Springer, 2011, pp. 584–587.

[21] A. Jylhä, Y.-T. Hsieh, V. Orso, S. Andolina, L. Gamberini, and G. Jacucci, "A wearable multimodal interface for exploring urban points of interest," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ser. ICMI '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 175–182. [Online]. Available: https://doi.org/10.1145/2818346.2820763

[22] A. Hub, J. Diepstraten, and T. Ertl, "Design and development of an indoor navigation and object identification system for the blind," in *ACM SIGACCESS Conference on Computers and Accessibility*, no. 77-78. ACM New York, NY, USA, 2003, pp. 147–152.

[23] F. Lang, A. Schmidt, and T. Machulla, "Mixed reality as assistive technology: Guidelines based on an assessment of residual functional vision in persons with low vision," in *International Conference on Computers Helping People with Special Needs*. Springer, 2022, pp. 484–493.

[24] K. Chaccour and G. Badr, "Novel indoor navigation system for visually impaired and blind people," in *International Conference on Applied Research in Computer Science and Engineering*. IEEE, 2015, pp. 1–5.

[25] Y.-T. Hsieh, A. Jylhä, V. Orso, S. Andolina, E. Hoggan, L. Gamberini, and G. Jacucci, "Developing hand-worn input and haptic support for real-world target finding," *Personal and Ubiquitous Computing*, vol. 23, pp. 117–132, 2019.

[26] Y.-T. Hsieh, V. Orso, S. Andolina, M. Canaveras, D. Cabral, A. Spagnolli, L. Gamberini, and G. Jacucci, "Interweaving visual and audio-haptic augmented reality for urban exploration," in *Proceedings of the 2018 Designing Interactive Systems Conference*, ser. DIS '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 215–226. [Online]. Available: https://doi.org/10.1145/3196709.3196733

[27] Y. Zhao, E. Kupferstein, B. V. Castro, S. Feiner, and S. Azenkot, "Designing AR visualizations to facilitate stair navigation for people with low vision," in *ACM Symposium on User Interface Software and Technology*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 387–402.

[28] I.-H. Hsieh, H.-C. Cheng, H.-H. Ke, H.-C. Chen, and W.-J. Wang, "Outdoor walking guide for the visually-impaired people based on semantic segmentation and depth map," in *2020 International Conference on Pervasive Artificial Intelligence (ICPAI)*. IEEE, 2020, pp. 144–147.

[29] R. Jafri, S. A. Ali, H. R. Arabnia, and S. Fatima, "Computer vision-based object recognition for the visually impaired in an indoors environment: a survey," *The Visual Computer*, no. 11, pp. 1197–1222, 2014.

[30] S. Gautam, K. S. Sivaraman, H. Muralidharan, and A. Baskar, "Vision system with audio feedback to assist visually impaired to grasp objects," *Procedia Computer Science*, vol. 58, pp. 387–394, 2015.

[31] K. Chen, Y. Huang, Y. Chen, H. Zhong, L. Lin, L. Wang, and K. Wu, "LiSee: A Headphone That Provides All-Day Assistance for Blind and Low-Vision Users to Reach Surrounding Objects," *ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 6, no. 3, pp. 1–30, 2022.

[32] G. Jocher, "YOLOv5 by Ultralytics (Version 7.0) [Software]." 2021.

[33] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Computer Vision – ECCV 2014*. Springer, Cham, 2014, pp. 740–755.

[34] F. J. Neyer, J. Felber, and C. Gebhardt, "Kurzskala Technikbereitschaft (TB, technology commitment)," *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*, 2016.

[35] J. R. Lewis, "The system usability scale: Past, present, and future," *International Journal of Human–Computer Interaction*, vol. 34, no. 7, pp. 577–590, 2018.

[36] B. Laugwitz, T. Held, and M. Schrepp, "Construction and Evaluation of a User Experience Questionnaire," in *HCI and Usability for Education and Work*, Springer. Springer, Berlin, Heidelberg, 2008, pp. 63–76.

[37] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, pp. 139–183.

[38] A. Cao, K. K. Chintamani, A. K. Pandya, and R. D. Ellis, "NASA TLX: Software for assessing subjective mental workload," *Behavior research methods*, no. 1, pp. 113–117, 2009.

[39] "Fast facts about vision loss," 2022. [Online]. Available: https://www.cdc.gov/visionhealth/basics/ced/fastfacts.htm

[40] "Share of individuals with vision loss in selected european countries in 2020, by severity," 2020. [Online]. Available: https://www.statista.com/statistics/1200735/vision-loss-by-severity-in-europe/

[41] Y. Zhao, E. Kupferstein, D. Tal, and S. Azenkot, ""It Looks Beautiful but Scary": How Low Vision People Navigate Stairs and Other Surface

Level Changes," in *ACM SIGACCESS Conference on Computers and Accessibility*, ser. ASSETS '18, 2018, pp. 307–320.

[42] Y. Zhao, C. L. Bennett, H. Benko, E. Cutrell, C. Holz, M. R. Morris, and M. Sinclair, "Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18, 2018, pp. 1–14.