Take My Hand: Automated Hand-Based Spatial Guidance for People with Visual Impairment*

Adil Rahman
adil@virginia.edu
University of Virginia
Charlottesville, Virginia, USA

Md Aashikur Rahman Azim
azim@virginia.edu
University of Virginia
Charlottesville, Virginia, USA

Seongkook Heo
seongkook@virginia.edu
University of Virginia
Charlottesville, Virginia, USA

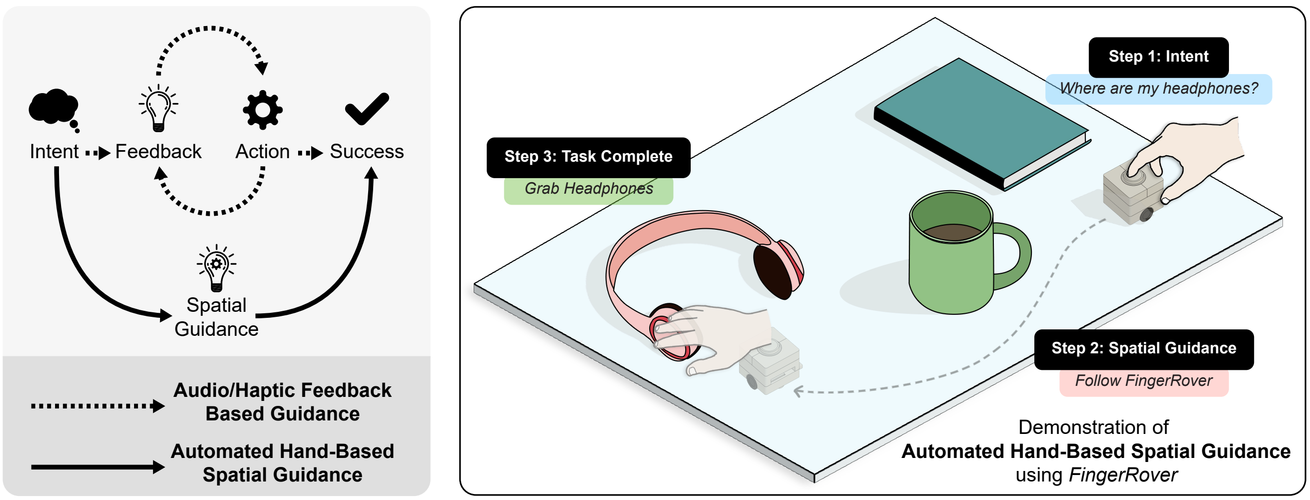

Figure 1: This paper presents automated hand-based spatial guidance, a technique that allows visually impaired users to reach targets on a surface using their hands without the need for interpreting directional cues. We facilitate this technique using FingerRover, an on-finger miniature robot.

ABSTRACT

Tasks that involve locating objects and then moving hands to those specific locations, such as using touchscreens or grabbing objects on a desk, are challenging for people with visual impairment. Over the years, audio guidance and haptic feedback have been a staple in hand navigation based assistive technologies. However, these methods require the user to interpret the generated directional cues and then manually perform the hand motions. In this paper, we present automated hand-based spatial guidance to bridge the gap between guidance and execution, allowing visually impaired users to move their hands between two points automatically, without any manual effort. We implement this concept through FingerRover, an on-finger miniature robot that carries the user's finger to target points. We demonstrate the potential applications that can benefit from automated hand-based spatial guidance. Our user study shows the potential of our technique in improving the interaction capabilities of people with visual impairments.

CCS CONCEPTS

• **Human-centered computing → Accessibility systems and tools**.

KEYWORDS

visual impairment; automated guidance; spatial guidance; miniature guiding robot; accessibility

ACM Reference Format:
Adil Rahman, Md Aashikur Rahman Azim, and Seongkook Heo. 2023. Take My Hand: Automated Hand-Based Spatial Guidance for People with Visual Impairment. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3544548.3581415

*This is the author's copy for the CHI 2023 paper *"Take My Hand: Automated Hand-Based Spatial Guidance for the Visually Impaired"*.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3581415

1 INTRODUCTION

Hand-based spatial navigation has been a long-standing and pervasive problem for people with visual impairments. What appears to be a simple task of picking up a phone from a desk involves heavy cognitive processing that we perform subconsciously - detecting the phone in the environment, estimating its location relative to us, planning the optimal path to reach the phone, and finally, moving the hand along the planned path to pick the phone up. The lack of visual modality makes performing even such simple tasks extremely challenging. More so, activities that require hand navigation to very specific targets, such as icons on a touchscreen or fields in a printed form, are particularly difficult to perform with precision.

Prior works have extensively explored the use of auditory [1, 16, 20, 29, 41, 51, 55] and haptic [8, 10, 12, 21, 22, 30, 40, 54] feedback to address this problem of spatial navigation for the visually impaired. A typical implementation of such systems involves the use of camera(s) running a computer vision pipeline to understand the user's surroundings, which is in turn used to track the target and then guide the user towards it using audio or haptic feedback. For example, StateLens [20] detects the user's finger position and targets on inaccessible touchscreens, and then guides the user to move their hands using repetitive directional speech cues (e.g., "Move left"). FingerReader [40], on the other hand, uses a finger-worn ring with multiple vibration motors to help the user keep their finger on the text. While these works attempt to solve some of the most critical challenges faced by the visually impaired community, they rely heavily on the user's ability to interpret and act upon the auditory and haptic feedback received from the system, requiring the user to manually move their hand to the target. This interpretation and execution of directional cues from the system involves extensive cognitive processing on the part of the user which can potentially increase their mental load [24, 28, 44, 58].

In this paper, we propose a novel approach, *automated hand-based spatial guidance*, which allows visually impaired users to reach targets on surfaces using their hands without the need for interpreting directional cues. Imagine a scenario where the visually impaired user orders a beverage on a touchscreen kiosk. However, instead of listening to any directional cues, the user simply places their hand on the touchscreen and it automatically moves towards the desired buttons to make the user's choice without the user doing anything manually. We believe that bridging the gap between guidance and execution by automatically moving the users' hand towards specific target points can potentially help visually impaired users conduct such everyday interactions that require locating and manipulating an object with less effort. To facilitate our idea of automated hand-based spatial guidance, we developed FingerRover, an on-finger, 2-wheeled miniature robot. FingerRover can locate itself and the target in three-dimensional space using a smartphone camera running an augmented reality pipeline, and then moves towards the target on a surface once the user places their finger on it, carrying along the user's finger (Figure 1). FingerRover also affords the capability to perform "risk-free exploration" [27] on touchscreens through a tapper that can register controlled touches. We present potential application scenarios that utilize these capabilities to demonstrate how automated hand-based spatial guidance can support people with visual impairments to perform everyday tasks with ease.

We conducted a study with 7 visually impaired users to validate the concept of automated hand-based spatial guidance through four application scenarios, while keeping audio feedback-based guidance as a baseline. The study results demonstrated the ability of automated hand-based spatial guidance to facilitate interaction for visually impaired users. Participants completed the tasks involving ordering items on a touchscreen kiosk and locating objects on a desk faster using our technique than using audio guidance, and reported an overall positive experience using the technique.

We summarize our contributions as follows:

(1) The concept of automated hand-based spatial guidance as an alternative modality to support visual-assistive interactions for people with visual impairments, and its demonstration using FingerRover, a miniature robot.
(2) A user study with visually impaired users that validates the effectiveness of automated hand-based spatial guidance in facilitating visual-assistive interactions.
(3) Application scenarios that demonstrate the potential ways in which automated hand-based spatial guidance can support people with visual impairments.

2 RELATED WORK

Our work draws inspiration from studies on hand guidance for the visually impaired. The lack of visual modality causes a considerable amount of challenges in sensemaking and spatial navigation. Previous works have attempted to bridge this gap of visual information by using various sensors and substitute modalities. Audio and haptic feedback methods have been extensively explored for various hand guidance applications.

2.1 Audio Feedback

Operating inaccessible interfaces is a serious challenge for the visually impaired. Audio feedback has been explored as an inexpensive means to assist visually impaired users to operate such interfaces. Feiz et al. [16] used dynamically generated verbal audio instructions to guide visually impaired users to independently navigate and write on printed forms. Similarly, LightWrite [55] used voice-based descriptive instructions to accurately teach visually impaired users to write English letters and numbers. Thakoor et al. [51] employed verbal directional cues to help visually impaired users orient their head-mounted camera towards objects of interest. Once oriented, their system would allow the user to localize and grasp the object. Crowd-sourcing has been leveraged to identify and label such inaccessible static [7, 19] and dynamic interfaces [20] in the wild. These applications use verbal directional feedback to provide hand guidance to visually impaired users for operating the inaccessible user interface based on the crowd-sourced information. However, while the audio channel can readily support information-rich feedback, the nature of such guidance may limit its applicability in noisy or context-sensitive settings [44].

2.2 Haptic Feedback

Haptic feedback has also been studied for accessibility technologies as it can be applied to various parts of the body as an alternative modality for conveying information. Unlike audio feedback, haptic feedback serves as a more discreet communication channel. Previous research has demonstrated a wide range of applicability of haptic feedback in assistive technologies, ranging from navigation [8] to reading printed text [40]. Hong et al. [21, 22] used haptic wristband with multiple configurations of vibrotactile motors to provide directional hand guidance to visually impaired users for

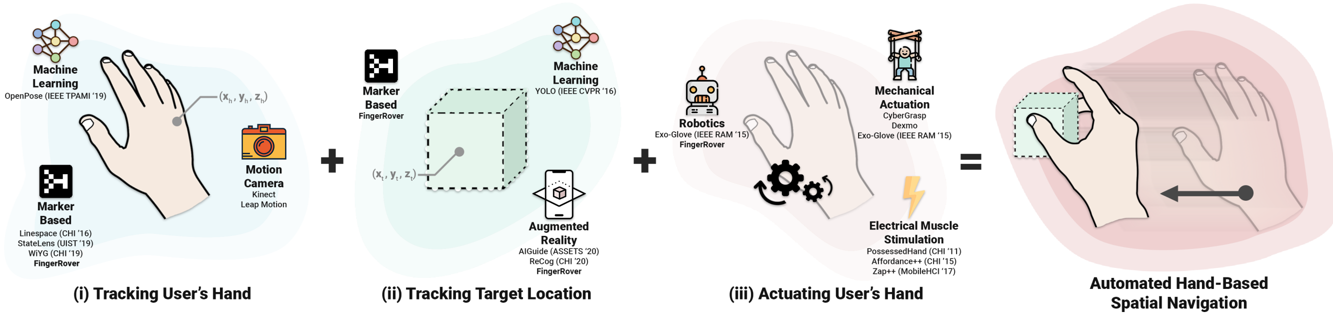

Figure 2: The components of automated hand-based spatial guidance.

target finding and path tracing on a 2D surface. FingerReader [40] and HandSight [44] used directional cues generated by finger-worn vibrotactile actuators to enable visually impaired users to read text off printed materials with a better understanding of their spatial layout. PalmSight [56] and FingerSight [23] translated visual cues to haptic feedback for use in navigation, localization and interaction with objects in the users' surroundings. A study by Shih et al. [39] showed that varying the intensity of vibrotactile feedback based on the distance and direction of the target object can help visually impaired users grasp such physical objects without knocking them down. A study by Huppert et al. [24] could be the most relevant to our approach. This study proposed to use a quadcopter drone to generate the directional force feedback by pulling a cable tethered to the user's finger. Their study showed that the drone-based hand guidance could help blind and visually impaired participants complete object localization tasks faster and more accurately than using audio feedback. While promising, the study did not investigate the diverse types of movements required to support accurate automated hand-based spatial guidance. Most haptic feedback based guidance techniques typically take the form of vibrotactile feedback. However, these types of feedback are less specific as they do not directly translate into movement and often require interpretation on the user's end. Moreover, several studies have shown that repeated application of vibrotactile feedback could lead to desensitization over time [6, 21, 44].

2.3 Audio + Haptic Feedback

Audio and haptic feedback often complement each other in assistive technologies to avoid overloading either sensory channel. PantoGuide [10] provided both audio and haptic guidance to users while exploring a tactile graph. Chase et al [10] emphasized the importance of tightly coordinating haptic and audio cues to maintain them in sync, and also the need to ensure that the guidance cues are perceptible under high mental load. AIGuide [52] exploits both audio and haptic feedback to guide visually impaired people in hand navigation applications. The user study showed a strong user preference towards guidance cues from multiple modalities, as opposed to single-modality feedback. Third eye [59], a shopping assistant for the visually impaired, uses audio feedback to help users ensure the wearable camera can view objects on the rack, while 4 vibration motors are embedded in the users' glove to provide vibrotactile directional cues for hand motion. Audio-tactile feedback has also been used to facilitate blind exploration of map interfaces [5, 14]. BotMap [14] is an actuated tangible tabletop interface that allows visually impaired users to interact with a map using pinch and zoom gestures. Robots scattered across the tabletop dynamically adjust their positions to represent different landmarks on the map, and users can explore the map using their tactile sense and system-generated audio cues. Chung et al. [11] explored the efficiency of audio and haptic feedback individually, and in combination, for assisting visually impaired users to locate targets in a three-dimensional space. Their results suggested audio feedback was the most efficient when used independently, or in conjunction with haptic feedback. While extensive research has explored the use of audio and haptic feedback to guide visually impaired users' hands through various application scenarios, the interpretation and execution of the generated directional cues is still left to the users. This could add to the cognitive overhead of the user, who may or may not interpret the cues correctly [24, 28, 44, 58].

3 AUTOMATED HAND-BASED SPATIAL GUIDANCE

We define the concept of *automated hand-based spatial guidance* as actuating the user's hand to move from one point in space to another without the user's interpretation of the guidance and manual effort. On a higher level, we argue that automated hand-based spatial guidance can be achieved using three core components: (i) tracking the user's hand, (ii) tracking the target location, and (iii) actuating the hand to move towards the target location (Figure 2).

Individually, these three components have been explored extensively by previous works, and they can be implemented in various ways. Hand tracking, for example, can be achieved using markers [16, 20, 46] or machine learning [9, 57] based methods, or by using tracking cameras such *Kinect* [32] or *Leap Motion* [53]. Similarly, target locations, depending upon the definition of *"target"*, can be tracked through augmented reality [2, 52] or machine learning [35] based methods. Techniques such as robotics [24, 25], electrical muscle stimulation (EMS) [15, 31, 47], and mechanical actuation [37] have been explored for actuating hand movement.

Depending upon the context, previous solutions in accessibility have tracked the user's hand and specific *target* points in conjunction to solve some critical challenges faced by the visually impaired. However, executing the actual motion towards the target has been left entirely to the user, who moves their hands relying solely on some form of system-generated guidance. In this paper, we combine all three components to demonstrate automated hand-based spatial guidance in the context of accessibility for the visually impaired.

While this concept can be implemented in several different ways, we envisioned our design to be portable, unobtrusive, and easy to put on or take off. Based on these design criteria, we developed *FingerRover*, an on-finger, 2-wheeled miniature robot capable of automated hand-based spatial guidance over a plane. FingerRover utilizes an augmented reality framework to continuously locate itself and its target in three-dimensional physical space through a paired smartphone and server system, and then moves towards the target location, carrying the user's finger with it.

4 FINGERROVER

In this section, we discuss the design and implementation of the hardware, tracking mechanism, and control system of FingerRover. Our FingerRover prototype is illustrated in Figure 3a.

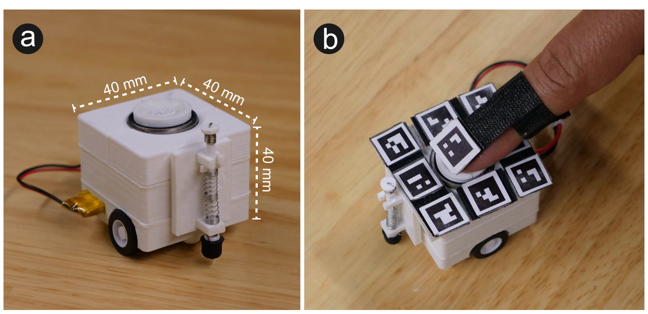

Figure 3: (a) FingerRover prototype, (b) marker configuration.

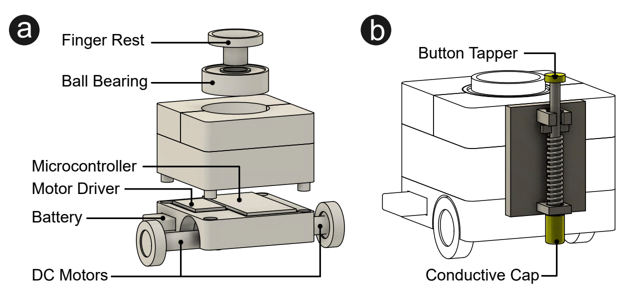

Figure 4: (a) FingerRover hardware design, (b) touchscreen tapper add-on.

4.1 Hardware Design

FingerRover is a lightweight robot weighing 52.85 grams. It is driven by two 14 mm × 4.5 mm wheels, each connected to a sub-micro DC planetary gear motor (Pololu #2359, 6D×21L mm) with a high gear ratio of 700 : 1 to generate a greater torque at the expense of speed. The DC motors are connected to a dual motor driver (DRV8835). FingerRover uses an ESP32 microcontroller (TinyPICO) which comes with an integrated Bluetooth module for wireless communication. A LiPo battery (100 mAh, 3.7 V) powers all the components of FingerRover. The maximum linear and angular speeds of FingerRover are measured to be approximately 37 mm/s and 106 deg/s, respectively. Under a standard operating load of 1 N applied normally, FingerRover is able to generate a maximum pulling force of 0.46 N.

All the components of FingerRover are enclosed within a 3D-printed cubic body of side 40 mm. A permanently lubricated stainless steel ball bearing (22 mm housing diameter, 8 mm shaft diameter) is used to allow FingerRover to rotate freely about its axis while the user rests their finger on top of it during operation. The finger-rest is 3D-printed and designed to fit in the ball bearing's shaft. The battery is housed in an external slot on the base of FingerRover and uses a 2-pin JST connector to allow for easy recharging. The components for building the prototype cost approximately 60 USD. Figure 4a illustrates the hardware design of FingerRover.

Lastly, FingerRover supports the capability to register touch on capacitive touchscreens through a tapper add-on attached externally to FingerRover's body. The tapper is constructed using a spring-based mechanism with a conductive cap on the bottom. The top end of the tapper is connected to the cap at the bottom through a conductive wire, which activates the cap only when the user touches the tapper, thus preventing accidental touches. The touchscreen add-on is illustrated in Figure 4b.

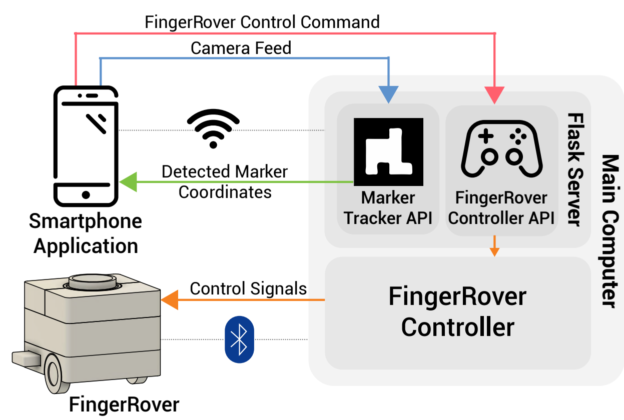

Figure 5: FingerRover control system.

4.2 Tracking Mechanism and Control System

FingerRover's tracking mechanism utilizes an augmented reality pipeline capable of recording the object coordinates in three-dimensional space and actively tracking it even when the object moves beyond the camera's field of view. We implemented this system as a mobile application in Unity using the *AR Foundation* framework [50]. We used a set of *ArUco* markers [17] for tracking FingerRover and the objects of interest in our implementation. An additional marker was attached to the user's finger (Figure 3b) to detect the user's hand location. The markers' corner points on the camera's viewport were tracked through a central *Flask* server using *OpenCV* at a rate of 15 frames/second. We used this to calculate the real-world coordinates and orientation of the markers by casting a ray through the obtained corner points on the camera's viewport and then tracking the feature points at which the rays intersect. FingerRover uses an 8-marker setup (Figure 3b) to reduce the chances of occlusion of the markers by the user's finger.

The central server serves as the main controller for the entire system. FingerRover is connected to the main computer running the server via Bluetooth, whereas the smartphone running the augmented reality pipeline is connected to the central server via WiFi. FingerRover is autonomously controlled by the main computer using convenient API endpoints provided by the central server – once FingerRover and the required target point are tracked by the augmented reality pipeline, their position and orientation differences are calculated and passed into a simple closed-loop control system such that the distance between them is minimized [30]. The control system is illustrated in Figure 5.

5 STUDY DESIGN

We designed a study to validate the concept of automated hand-based spatial guidance by evaluating its effectiveness in allowing visually impaired users to perform various everyday tasks and to gain an insight into their experiences of using the automated hand-based spatial guidance and the audio feedback based guidance.

5.1 Task Description

To evaluate the efficacy of automated hand-based spatial guidance in supporting a diverse range of applications for the visually impaired, we chose four common accessibility problems as our tasks – operating a touchscreen, locating objects on a desk, assembling an object, and interpreting shapes traced by hand movement.

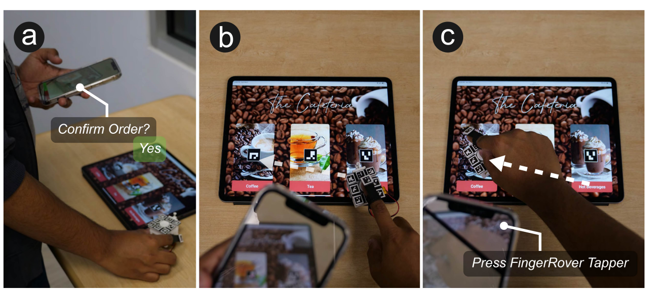

Figure 6: Operating touchscreen interface using FingerRover. (a) User verbally chooses a beverage, (b) FingerRover moves towards appropriate buttons, (c) user makes selection on the touchscreen using the tapper add-on.

Task #1: Operating a Touchscreen. In this task, the participants operated an inaccessible touchscreen interface running a simulated beverage-ordering kiosk (Figure 6). The participants first used voice commands to pre-select their beverage of choice and additional properties such as beverage size and strength [20]. After confirming their choice, the participants were guided by the guidance system to make the appropriate selection on the touchscreen kiosk. A total of 5 selections were required to order a beverage – *category*, *beverage*, *strength*, *size*, and *order confirmation*. In the case of an incorrect selection, the participants were guided to undo the mistake using the ongoing method. Participants were required to order two beverages using each method to complete this task.

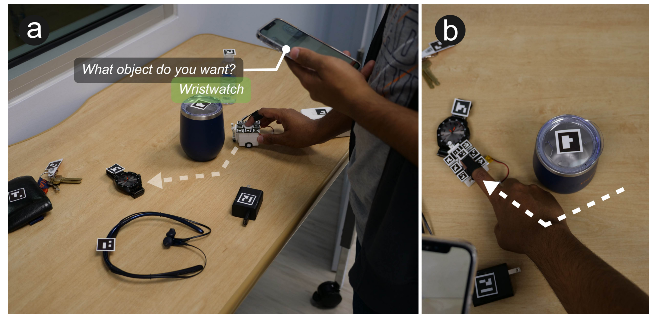

Figure 7: Locating objects using FingerRover. (a) User verbally chooses an object, (b) FingerRover moves towards the object.

Task #2: Locating Objects. In this task, the participants were required to locate objects of interest from a group of 8 everyday objects placed randomly on a desk (Figure 7). The participants were first presented with a list of objects on the desk, following which they used a voice command to select an object that they wanted to locate. After making a choice, the participants received audio/spatial guidance to move their hands to locate the object. Participants were required to locate three objects using each method to complete this task.

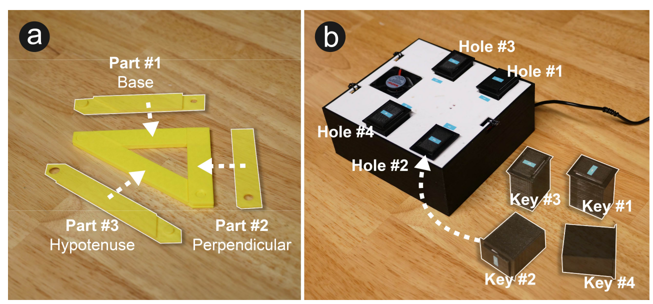

Figure 8: The two objects used for object assembly. (a) Right-angled triangle, (b) custom fan.

Task #3: Assembling an Object. In this task, the participants were required to assemble a custom 3D-printed object with the help of remote assistance provided by the experimenter (Figure 8). The parts of the 3D-printed object were provided to the participants as a disassembled product, and the experimenter played the role of a remotely located customer service agent assisting the object assembly. The experimenter used both the audio and spatial guidance methods to guide the participants' hand to each part of the object, sequentially, in the order in which the object needed to be assembled. During the audio guidance method, the experimenter

Table 1: Statements used for qualitative system evaluation.

Task	Statement
Operating Touch Screen	**[Low Effort]** I was able to operate the touch screen without much effort. **[Operating Confidence]** I felt confident operating the touch screen. **[Making Selection]** I was confident about making a selection on the screen.
Locating Objects	**[Easy to Locate]** It was easy to locate the desired object. **[Locating Confidence]** I felt confident locating the objects. **[Avoiding Obstacles]** The technique allowed me to easily navigate through obstacles.
Assembling an Object	**[Liked Assistance]** I liked the assistance that I received to assemble the object. **[Assembling Confidence]** I felt confident using the technique to assemble the object.
Shape Sensemaking	**[Shape Understanding]** I was able to clearly understand the shapes through FingerRover's movement.
Overall Interaction	**[Technique Effectiveness]** The technique was effective in guiding my hands towards the targets. **[Task Confidence]** I felt confident using the technique to perform the tasks. **[Ease of Use]** I was able to easily guide my hands in the direction specified by the technique. **[Physical Fatigue]** Using this technique to complete the tasks involved considerable physical fatigue. **[Mental Fatigue]** Using this technique to complete the tasks involved considerable mental fatigue.

talked to the participants to inform them about their environment and provided verbal directional cues to reach the required parts of the object. For the spatial guidance method, the experimenter used the FingerRover in conjunction with audio guidance technique, with the experimenter explicitly controlling FingerRover to provide spatial assistance to locate points of interest (such as object pieces and assembly locations), and verbally conversing with the participants to provide additional details about the environment and assembly instructions. This task was performed only once with each method.

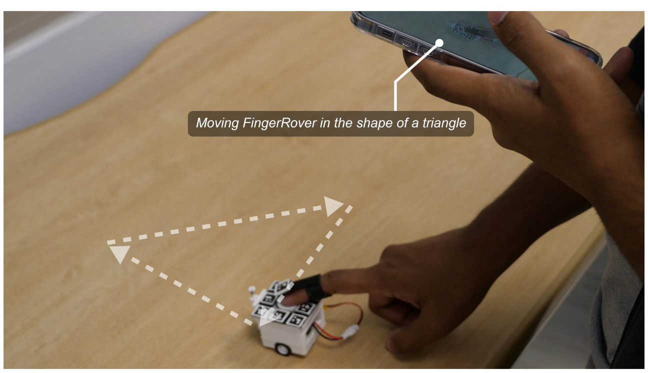

Figure 9: FingerRover tracing a triangle through its movement.

Task #4: Shape Sensemaking. This task was exclusively designed for the spatial guidance method to evaluate its ability to convey shape information via automated hand movement. In this task, the FingerRover traced simple shapes and the participants were asked to guess the shapes (Figure 9). This task was repeated three times, each time with a different shape – square, triangle, and diamond.

5.2 Apparatus

We facilitated automated hand-based spatial guidance using our *FingerRover* prototype. To allow for a longer runtime during the study, we used a larger 900 mAh battery to power FingerRover, which added approximately 12 grams to the total weight of the device. A custom smartphone application integrated with text-to-speech (TTS) support was used to facilitate audio feedback based guidance. Both FingerRover and the TTS application were connected to a *Flask* server with a custom desktop interface, which acted as the primary control system. We simulated a coffee machine interface on a 12.9-inch iPad Pro for the task of operating a touchscreen. For the task of assembling an object, we designed two custom 3D-printed objects consisting of multiple parts.

5.3 Procedure

The study was designed to be conducted indoors in a controlled lab environment. Participants were seated during the entire duration of the study. They were first briefed about the study and their demographic information was recorded. The participants were then given a short tutorial to ensure that they were familiar with both spatial and audio guidance methods. For both methods, the participants were required to hold the smartphone running the TTS application in their non-dominant hands, while the dominant hand was used to perform the tasks following the spatial and audio guidance. The TTS application served as a conversational agent controlled by the experimenter and was used by the participant to make choices involved in certain tasks, such as choosing a beverage on the coffee machine interface [20]. Once the participant verbally conveyed their choice to the conversational agent, they were guided by each method towards physically executing those choices. For spatial guidance, FingerRover guided the participants' finger physically towards the points of interest. The audio guidance comprised of 4 directional cues for moving the hand *forward*, *backward*, *left*, and *right*, with 2 additional cues for *lifting* and *lowering* the hand. For both methods, the system provided voice feedback *grab object* or *press button* when the participants' hand was near the target. Following the tutorial, the participants were asked to perform each of the tasks using both methods. After completing each task with a method, the participants were required to fill out a questionnaire using 7-point Likert scale and provide feedback about their experience of using each method for the task. Once all the tasks were completed, the participants took a final survey reporting their overall experience of using both methods.

All the tasks were performed in the same order with alternating methods. The order of methods was counterbalanced to minimize the ordering effect. Thus, half of the participants performed each task first using spatial guidance, and then using audio guidance, whereas the other half performed each task first using audio guidance, and then using spatial guidance. To gain insight into the efficiency of each method, the completion time was recorded for each task (with the exception of shape sensemaking) once the participants had conveyed their choice to the conversational agent and the audio/spatial guidance was initiated. However, to ensure that the tasks were not rushed at the expense of accuracy, the participants were not guided to complete the tasks as quickly as possible.

Since the goal of this study was to evaluate the effectiveness of automated hand-based spatial guidance, we replaced the augmented-reality based tracking system of FingerRover with a *Wizard-of-Oz* approach. The experimenter was thus in charge of visually tracking the participants' hand and target location, and controlling the FingerRover to accurately guide the participants' hand to the target. The same technique was used for facilitating audio feedback based guidance. This ensured that we were able to evaluate both methods at perfect tracking conditions, thus eliminating any confounding bias associated with camera positioning and marker tracking errors. Since a Wizard-of-Oz technique was used, participants did not wear the tracking markers on their fingers.

6 PRELIMINARY STUDY

Before evaluating our system with the target population of visually impaired individuals, we first conducted an IRB-approved preliminary study with blindfolded participants without any visual impairments to validate FingerRover's ability in facilitating automated hand-based spatial guidance, and to ensure that both the spatial and audio guidance methods were capable of supporting the proposed set of tasks in the user study.

6.1 Participants

We recruited 12 participants (4 female, 8 male) without any visual impairments from our university using email groups and word of mouth. The participants' ages ranged between 21 and 36 years old, with an average age of 27.8 years ($\sigma = 4.28$). All participants were right-handed. The experiment sessions took one hour per participant to complete, and each participant was compensated with $20 USD for their time.

6.2 Additional Procedure Details

The preliminary study followed the same procedure as described in the study design. Since the participants did not have any visual impairment, they were blindfolded before performing each task. For the tasks of operating a touchscreen and locating objects on a desk, the participants were given the freedom to pre-select their choices of beverages and objects, respectively. For the object assembly task, we used a 3D-printed right-angled triangle (Figure 8a), which could be deconstructed into 3 unique components – *perpendicular*, *base*, and *hypotenuse*, each with a uniquely sized hole-plug pair to facilitate connection with the other parts.

6.3 Results

All participants were able to successfully complete the tasks using both methods, thus validating the feasibility of our study design and FingerRover's ability to facilitate automated hand-based spatial guidance. The task completion time for each task is illustrated in Figure 10. Participants' responses to the survey questionnaire are provided in the appendix. Overall, the participants provided positive feedback for both methods, but reported their preference towards spatial guidance. We further discuss our findings from each interaction below.

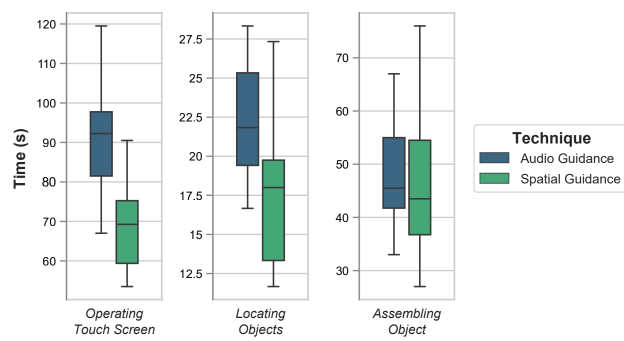

Figure 10: Task completion time for the preliminary study.

6.3.1 Operating a Touchscreen. On average, the participants completed this task approximately 23% (21 seconds) faster using spatial guidance as compared to audio guidance. A repeated measures ANOVA showed that the difference was statistically significant ($F(1, 11) = 30.96, p < 0.001$). An incorrect selection was made during one of the trials using audio guidance. However, no incorrect selections were made while using FingerRover. Participants' responses for this task indicated an overall preference towards spatial guidance, with participants reporting that FingerRover allowed them to operate the touchscreen interface with much less effort and higher confidence as compared to the audio guidance method.

6.3.2 Locating Objects. Participants completed this task approximately 21% (5 seconds) faster on average using spatial guidance as compared to audio guidance. A repeated measures ANOVA showed a significant effect of method on task completion time ($F(1, 11) = 5.51, p < 0.05$). Participants' responses for both methods in this task were positive, with a slight preference towards FingerRover.

6.3.3 Assembling an Object. To complete this task, both the methods took approximately the same amount of time, with spatial guidance being only 3% (1 second faster) on average. A repeated measures ANOVA did not find a significant effect of method on task completion time. In the Likert scale responses, participants stated that they liked both methods equally, but felt more confident assembling the object using the audio guidance method. This could be attributed to the fact that during the spatial guidance method, FingerRover could only help the participants fetch the three parts sequentially, but the act of assembling the object required explicit verbal guidance.

6.3.4 Shape Sensemaking. Participants reported being able to clearly understand the shape conveyed by the FingerRover's movement. All the participants could correctly interpret the square and triangle shapes. Four participants were able to precisely guess the diamond shape, whereas six participants guessed it to be a rhombus. Two participants interpreted the diamond shape as a trapezium and a rectangle.

6.4 Discussion

The results show the potential benefits of the spatial guidance technique over audio guidance, and the participants' responses resonated with our expected advantages that the automated spatial guidance technique can offer over feedback-based guidance techniques. Participants found FingerRover to be more effective in guiding their hands towards the targets and reported having lower physical and mental fatigue completing the tasks using FingerRover. For the spatial guidance technique, participants remarked that *"the best part was I didn't need to think about the directions and I was easily finding everything just placing my finger on the FingerRover"* (P1) and that *"following the rover was much easier than trying to interpret commands"* (P5). Furthermore, participants pointed out some drawbacks of the audio guidance method – *"I was getting a bit distracted because I was not sure if I was moving my hands properly"* (P1), *"I was not sure how far I should move my hand"* (P11).

Our findings from the preliminary study motivated further evaluation of our system's usability with the target demography of visually impaired users. However, the preliminary study also highlighted some flaws in our study design. The preliminary study revealed an issue with the choice of object used for the object assembly task, since spatial guidance was only applicable for locating the different components sequentially, but did not support the actual assembling of the object. To address this, we designed an alternative object for the task – a custom fan consisting of four identical removable keys ($44\ mm \times 34\ mm \times 24\ mm$) which, when placed in specific positions following a specific sequence, could power the fan (Figure 8b). We believe this design would be more representative for evaluating the task since the spatial guidance technique could be used to not only locate the parts, but also locate specific positions where these parts had to be placed. Additionally, we realized that asking participants to choose beverages and objects themselves led to some degree of fatigue. Thus, to prevent fatigue associated with making choices, we decided to randomly assign the beverages and objects to the participants that they would need to choose in the tasks of operating a touchscreen and locating objects on a desk, respectively.

7 USER STUDY WITH VISUALLY IMPAIRED USERS

We conducted a user study (UVA IRB-SBS #4591) with visually impaired participants to gain insight into their interaction experiences while using automated hand-based spatial guidance.

7.1 Participants

We recruited 7 participants (4 female, 3 male) from local communities for the visually impaired through newsletters, social media posts, and word of mouth. Three participants were completely blind, while four participants had low vision, three of whom were categorized as legally blind. All participants reported having trouble operating touchscreens, locating things, and assembling objects in their daily lives. The participants' ages ranged between 30 and 74 years old, with an average age of 51.9 years ($\sigma = 15.3$). All participants were right-handed. Each participant was compensated with a $50 USD gift card for their time and travel expenses. Additional details regarding the participant demography are included in Table 2.

7.2 Additional Procedure Details

This study included the changes which were identified in the findings of our preliminary study. Additionally, prior to the study, we interviewed the participants to understand the situations where they faced challenges in tasks that required hand manipulation. Based on their input, we staged the study settings to mimic such situations to test if automated hand-based spatial guidance can be effective at supporting the participants in those scenarios. Thus, two participants (P4, P5) were requested to remove visual aids, and for two participants (P5, P6), the studies were conducted in dim light conditions. Finally, at the end of the study, we conducted a semi-structured interview to gain additional insights about their interaction experience.

7.3 Results

The participants' responses to the survey questionnaires and the task completion time for each task are illustrated in Figure 11. We report our findings from the study in terms of the *efficiency*, *accuracy*, and *usability* of automated hand-based spatial guidance, while keeping audio guidance as our baseline.

7.3.1 Time and Accuracy. The participants were able to successfully complete all the tasks using both the methods, and the entire study lasted for approximately 75 minutes on average. We now discuss the time and accuracy results for each task below.

- **Operating Touchscreen:** Participants completed this task approximately 24% (17 seconds) faster using spatial guidance ($\mu = 53.1s,\ \sigma = 9.1s$) as compared to audio guidance ($\mu = 69.8s,\ \sigma = 23.3s$). A higher difference was observed for completely blind participants (P1, P2, P7) who completed this task approximately 30% (23 seconds) faster on average using spatial guidance ($\mu = 55.2s,\ \sigma = 11.8s$) as compared to audio guidance ($\mu = 78.5s,\ \sigma = 16s$). Participants with low vision completed this task approximately 19% (12 seconds) faster on average using spatial guidance ($\mu = 51.5s,\ \sigma = 7s$) as compared to audio guidance ($\mu = 63.25s,\ \sigma = 26.6s$). A repeated

Table 2: Participant demography of our user evaluation with 7 visually impaired users.

ID	Age/Sex	Vision Level	Diagnosis	Additional Details
P1	58 F	Blind	Retinitis Pigmentosa	Blind since 2012.
P2	39 M	Blind	Retinopathy of Prematurity	Blind since birth.
P3	30 M	Legally Blind	Bardet-Biedl Syndrome	Legally blind since 2014.
P4	53 F	Legally Blind	Retinitis Pigmentosa	Legally blind since 2011. No peripheral vision.
P5	65 F	Low Vision (20/50)	Diabetic Retinopathy, Cataract	Blind right eye since 2010.
P6	74 M	Legally Blind	Macular Degeneration, Glaucoma	Legally blind since 2016.
P7	44 F	Blind	Retinal Detachment	Condition since 2011. Some light perception.

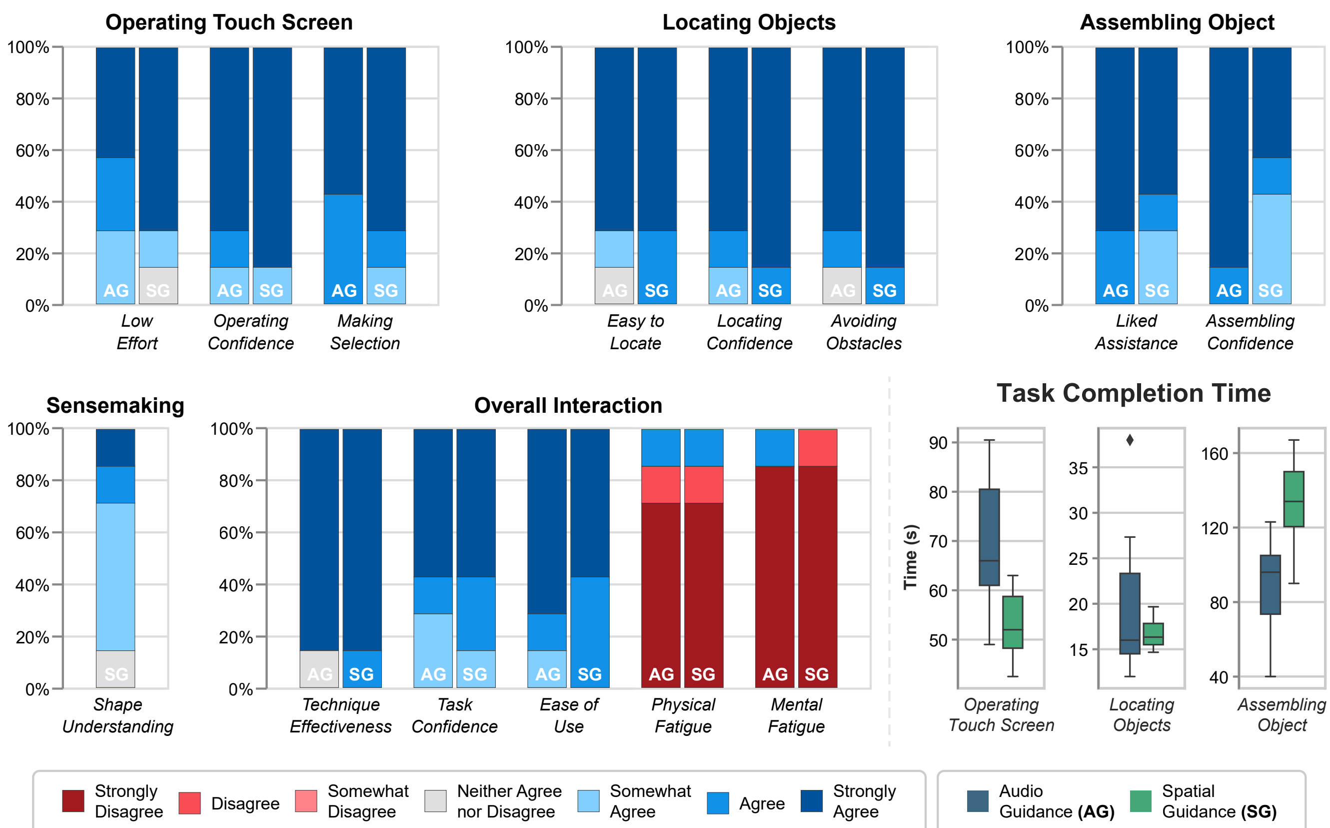

Figure 11: Visually impaired participants' qualitative assessment and completion time for each interaction.

measures ANOVA for all the participants revealed a statistical significance between the guidance method and task completion time ($F(1, 6) = 7.35$, $p < 0.05$). Furthermore, we observed a total of 25 mistouches on the touchscreen during trials with audio guidance, 5 of which led to incorrect selections on the interface. 20 of the total mistouches were registered by completely blind participants. Only 1 mistouch was observed during a trial using spatial guidance (P1), which also led to an incorrect selection.

- **Locating Objects:** Participants completed this task approximately 17% (3 seconds) faster on average using spatial guidance ($\mu = 16.8s$, $\sigma = 4.3s$) as compared to audio guidance ($\mu = 20.2s$, $\sigma = 11.2s$). However, a repeated measures ANOVA did not find a significant effect of method on task completion time ($F(1, 6) = 1.32$, $p = 0.29$).
- **Assembling an Object:** Participants found audio guidance to be more efficient for this task. Participants were approximately 34% (45 seconds) faster on average using audio guidance ($\mu = 88s$, $\sigma = 28.8s$) as compared to spatial guidance ($\mu = 133.1s$, $\sigma = 26.1s$). Guidance method was found to have a statistically significant effect on the task completion time using a repeated measures ANOVA ($F(1, 6) = 10.97$, $p < 0.05$).

- **Shape Sensemaking:** All the participants were able to accurately recognize the triangle shape. Five participants were able to recognize the square shape, with one participant (P3) interpreting it as a trapezoid, and another (P1) misclassifying it as a hexagon. One participant (P2) interpreted the diamond/rhombus shape as an *off-angled square*, whereas another (P4) interpreted it as a type of parallelogram. Two participants (P1, P7) classified the diamond as a square, whereas three (P3, P5, P6) classified it as a rectangle.

7.3.2 Usability. All participants reported a positive experience for both spatial and audio guidance methods in the survey questionnaires and the post-study interview. All participants had prior experience of using some kind of audio-feedback based assistive technology, such as *Aira*, *Be My Eyes*, *Siri*, and even *FaceTiming* family members, to perform various daily activities such as locating objects. However, participants' responses indicated an overall preference towards spatial guidance during the study, and they reported that it was easy to get familiar with the technique within a few trials. Two participants (P4, P5) stated that they preferred spatial guidance over audio since it was easier for them to follow the robot instead of interpreting directional cues from the audio, with one participant reporting that following audio guidance led to significant mental exhaustion – *"What I found difficult was I really had to use good listening skills in it, and so I did get mentally tired from having to really concentrate. [The robot] took off some of the pressure to listen as intensely as I had to for just the audio"*, (P5). Two participants also appreciated the tangible nature of the spatial guidance technique (P2, P7) – *"Sometimes if you want to reach out for something, it would be nice to have something physical, something we can touch and feel. [Audio Guidance] is just talking to me, it's still just a voice, but the robot is a machine, it's physical"*, (P7). On the prospect of integration of spatial guidance to other visual assistive technologies, one participant compared spatial guidance to *VoiceOver* – *"With VoiceOver, what people do is they get faster and faster on the voice, where a normal person can't even understand what's being said. But you still have to go through all that reading to get to something, and it's very laborious, say, if you're in an administrative scenario. I can see [spatial guidance] cut through a lot of that. It would be superior to VoiceOver, a vast improvement of what the standard is now"*, (P6). We now discuss the user experience for each task individually.

- **Operating Touchscreen:** Participants found spatial guidance effective in helping them operate the inaccessible touchscreen. Participant responses from the survey questionnaire and post-study interview suggest that spatial guidance required low effort to operate the touchscreen, and that they were confident in using the technique to operate the touchscreen. Three participants (P1, P2, P7) expressed their concerns about positioning their hands during their trials with audio guidance – *"One thing that makes [audio guidance] challenging is that you have to move your finger through the air rather than along the surface of the screen, so you don't have the feedback of actually touching the screen and thus may not be aware of how far above the screen you are. There is also the possibility of accidental touch. A person who is not a very skilled user might accidentally touch with two fingers two different elements on the screen"*, (P2). The physicality of spatial guidance was reportedly able to avoid these problems – *"I liked that the robot gives you haptic and auditory feedback. So you actually have something that you can touch as it moves across the screen"*, (P2). Some low vision participants (P4, P5, P6) were able to vaguely perceive the shape of the buttons on the touchscreen interface using their residual vision. As a result, during trials with audio guidance, they were able to move their hands efficiently towards the next available button in the given direction. However, despite the advantage, they reported preferring spatial guidance for the interaction as they found verbal directional cues hard to get used to (P4) and mentally exhausting (P5). Some comments from the participants included – *"I loved it! It's very, very helpful. I could have done it with my eyes closed"*, (P5); *"[The interaction] was very simplistic, which is good. There was a natural flow, because the menu can change, but your actions remain the same"*, (P6).

- **Locating Objects:** Participants found spatial guidance to be effective in helping them locate objects on a desk. Most participants (P3, P4, P5, P6, P7) shared a common sentiment about FingerRover, stating that they have trouble locating items in their daily lives and that this technique could be very useful for them – *"I liked it very well, I would love to have this at home"*, (P6). There were no instances of object collision while locating objects using either audio or spatial guidance. However, three participants (P2, P3, P5) perceived spatial guidance to be better at preventing such instances of object collision – *"I feel [the robot] is a little safer because it will navigate around objects, as opposed to just the audio where I could knock something over to get to something else"*, (P5). One participant appreciated the precise guidance that FingerRover provides – *"A couple of objects were next to each other, so I could've certainly picked up the wrong thing [using audio guidance]. I think the robot was better than the audio. I was able to find everything that you asked me to find instantly. I didn't pick up the wrong things"*. However, one participant (P2), who was an O&M instructor, stated that he preferred performing a *manual grid search pattern* to find objects on a desk as opposed to using any technology.

- **Assembling an Object:** For this task, participants demonstrably preferred audio guidance over spatial guidance, even though the spatial guidance method for this task also featured verbal communication. This was observed because operating FingerRover restricted participants from using both their hands together for assembling the object, with the participants reserving one hand exclusively to follow FingerRover, and the other for holding the pieces to assemble the object. Participants were able to use both their hands freely during trials with audio guidance, using one hand for locating the correct piece, and the other for locating the positions on the object's body where the selected piece would go. Since there were only 4 moderately sized pieces that were required to assemble this object, participants were able to cycle through each piece easily. However, several participants (P2, P3, P5, P6) acknowledged the value that spatial guidance could bring for more complicated object assembly tasks featuring multiple small-sized pieces – *"It's easier to pick things up [using the robot] if they're in a jangled mess"*, (P3); *"These objects happened to*

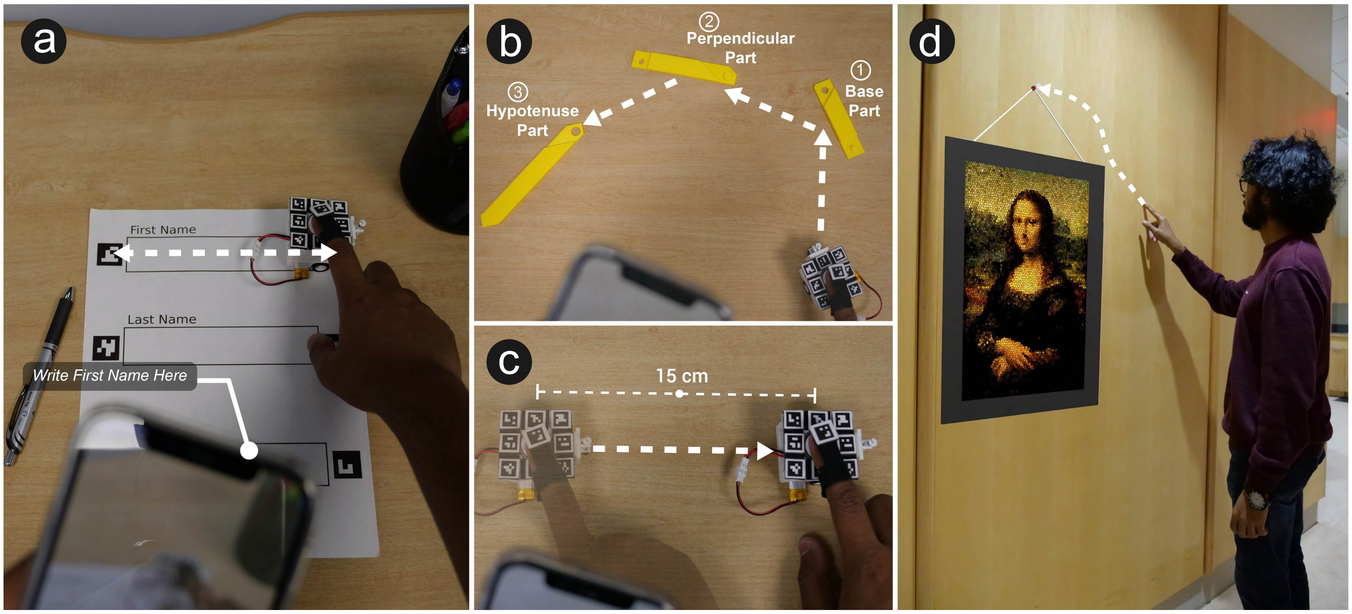

Figure 12: Some application scenarios supported by spatial guidance. (a) Writing on printed form, (b) sequentially fetching objects in an object assembly task, (c) taking physical measurements, and (d) locating point for drilling a hole.

be big enough for me to see, so the verbal [guidance] was just adequate. But if I am unable to see it, I would definitely need [the robot]", (P5). One participant (P1) commented that unclear verbal instructions could also lead to potential misunderstandings between the user and the guide, which could be avoided by using the robot as the guide would have explicit spatial controls.

- **Shape Sensemaking:** Participants demonstrated being able to interpret features such as the number of sides and angles of the shapes traced via FingerRover's movement. After one of the trials, a participant (P1) revealed that the direction in which a shape is drawn is significant for her, since that is how she had learned to associate shapes – *"I was slightly confused with the shape thing, because in my world, a triangle goes like this (. . .) [traces a triangle]. So, I had a little trouble orienting"*. However, despite the shape being traced in a direction different than what would be normal to her, she was able to correctly interpret two shapes out of a total of three. Another participant (P2) associated the interaction in this task with his prior experiences of learning sign language, suggesting further applicability of spatial guidance in other visual-assistive domains.

8 APPLICATION SCENARIOS

In this section, we present some application scenarios based on our ideas, prior research work, and discussions with visually impaired users from our user study to illustrate how automated hand-based spatial guidance can help people with visual impairments in various scenarios. We group the application scenarios into three categories – *everyday accessibility*, *remote assistance*, and *computer-aided affordances*.

8.1 Everyday Accessibility

8.1.1 Operating Touchscreen Interfaces. Modern smartphones come with visual-accessibility features that allow blind users to operate these devices. However, public touchscreen interfaces such as kiosks and in-flight entertainment systems do not support these accessibility features. In our study, all participants reported being unable to operate these interfaces without the help of their family members. Spatial guidance can effectively support visually impaired users in operating such inaccessible touchscreen interfaces independently, as was demonstrated in our user study (Figure 6).

8.1.2 Locating Objects. Locating objects in their surroundings is a common grievance among people with visual impairments, which was also reflected in the participants' (P3, P4, P5, P6, P7) responses during our study. Our user study demonstrated that not only was spatial guidance effective in assisting visually impaired users to locate objects on a desk, it also had the potential to prevent users from knocking things over accidentally by navigating around potential obstacles (Figure 7).

8.1.3 Independently Writing on Paper. The lack of visual modality can make it challenging for visually impaired users to independently write on paper. Spatial guidance can be integrated with preexisting assistive technologies (such as *WiYG [16]* to facilitate such interactions – the user can point the smartphone camera in the approximate direction of the printed form to locate all entry fields, and then spatial guidance could be employed to guide the users hand towards each entry field while the smartphone verbally announces the field name (Figure 12a).

8.2 Remote Assistance

8.2.1 Sensemaking in Remote Learning. Previous works have underlined the benefits of media synchronization in the domain of education [4, 26]. However, devices that can enable such interactions (e.g., PHANToM Omni [49]) are often expensive and bulky. There also exists tabletop interfaces such as F2T [18] which facilitate two-dimensional data exploration and sensemaking through the use of force-feedback, but their range of movement is often constrained to within the interface. Our results from the shape sensemaking task suggest that the spatial guidance afforded by FingerRover can be used to convey shape information, making it a low-cost, miniature alternative capable of supporting media synchronization. Consider a remote learning situation where a visually impaired student is learning physics, and the instructor wants to communicate the shape of a sine wave. The instructor can do so by controlling FingerRover to trace the shape of the sine wave, thus spatializing the visual information (Figure 9). The versatility of the spatial guidance can also allow the student to interactively change parameters, such as the frequency of the sine wave, and understand how it affects the waveform.

8.2.2 Assembling an Object. Assembling objects such as furniture often involves identifying the individual components and then combining them in a specific sequence. The non-standard nature of such tasks makes it challenging for the visually impaired. Most participants in our user study reported that they needed physical assistance from their family members to perform even trivial assembly tasks. Spatial guidance can allow visually impaired users to receive accurate remote assistance to help them assemble objects (Figure 12b). Furthermore, one participant (P3) also suggested using FingerRover in mechanical workspaces to assist in tasks such as drilling holes by accurately locating the point of interest (Figure 12d).

8.2.3 Guiding Physical Activities. During the post-study interview, one participant (P5) suggested the use of spatial guidance to assist visually impaired users' hand postures during physical exercises, such as *chair zumba*. Tethering the relative spatial position of one person's hand with that of another person could allow for several interesting applications, which would be even more relevant for visually impaired users. While the current implementation of FingerRover does not support three-dimensional spatial manipulation of the entire arm, spatial guidance using other hand actuation mechanisms, such as EMS [31] or *LineFORM* [33], can be considered in facilitating these types of interaction. P5 stated that not only would such technology promote *self-independence* among the visually impaired, but would also allow other people to feel more comfortable including them in activities, thereby promoting *inclusivity*.

8.3 Computer-Aided Affordances

8.3.1 Obstacle Avoidance. People with visual impairments run the risk of accidentally knocking over objects while trying to blindly navigate a surface, which was a common grievance of most participants in the study. The precise movements afforded by spatial guidance can allow the users to avoid this risk by navigating around obstacles to reach their desired target. Figure 7b illustrates one such scenario where FingerRover moves around the coffee mug to reach the wristwatch.

8.3.2 Taking Physical Measurements. Taking physical measurements is often challenging for the visually impaired. Spatial guidance can be exploited to accurately represent dimensions in the physical space. Figure 12c illustrates this interaction using FingerRover – the user places their finger on FingerRover and asks the smartphone to move it 15 cm towards the right; FingerRover moves approximately 15 cm towards the right of its starting position, thereby indicating the required measure to the user.

9 DISCUSSION

In this section, we discuss the additional benefits of using automated hand-based spatial guidance, supporting interactions with automated hand-based spatial guidance, limitations, and future research directions.

9.1 Benefits of Automated Hand-Based Spatial Guidance

The results from our user study demonstrated several benefits of using automated hand-based spatial guidance. Both blind and low-vision participants found the technique to be helpful. They were able to easily adapt to spatial guidance, and found it to be effective in operating a touchscreen and locating objects on a desk. These trends were also observed in our preliminary study with blindfolded users. Additionally, visually impaired participants acknowledged its utility in various other application scenarios. The benefits that we observed from our technique resembled those of prior research work that affords new interactions by explicitly manipulating the spatial modality [42, 48].

During the trials using audio guidance, we observed different participants had different spatial thresholds for moving their hands, where some participants made minute changes to their hand positions upon listening to directional cues, whereas others overshot the target points with drastic hand movements. Spatial guidance introduced uniformity to the process, since the participants' hand position was tethered to the spatial location of FingerRover.

FingerRover has a slow movement speed which directly influences the speed at which the users' hand moves. However, we observed significant task completion time benefits of using spatial guidance for the task of operating a touchscreen, even for participants with low vision. This could be attributed to the fact that despite being slow, FingerRover afforded accurate spatial guidance which prevented incorrect selections, thereby maintaining a higher overall efficiency. This suggests that spatial guidance could be potentially useful in interfaces where input accuracy is important and recovering from misselections is expensive, e.g., payment systems.

However, it is important to note that we do not suggest spatial guidance as a replacement for audio/haptic feedback based guidance techniques, and instead, we present spatial guidance as another modality that can be explored to facilitate accessibility for the visually impaired.

9.2 Supporting Interactions With Automated Hand-Based Spatial Guidance

Beyond FingerRover, it is exciting to think of the many possibilities in which the automated hand-based spatial guidance can be implemented. Several off-the-shelf actuated robots, such as *Toio* [43], exist which could offer various interactive capabilities. Past research has also explored the development of miniature robots [13, 45] which could be used to demonstrate automated hand-based spatial guidance. Unlike hand navigation methods that utilize tactually-encoded guides [40, 44], automated hand-based spatial guidance does not require interpretation of haptic signal to move the hand, although the participants can still kinesthetically feel the movement of the device. This frees up the tactile channel to be used to encode additional information such as the direction or distance to the target using braille displays or vibrotactile actuators. We consider electrical muscle stimulation (EMS) based methods to be a viable candidate for the implementation of our technique given its granular controls capable of enabling fine-grained controls involving multiple fingers [34]. EMS can also allow objects to communicate their affordances [31], which can be beneficial in tasks requiring dexterous manipulation such as assembling objects.

When designing interactions utilizing automated hand-based spatial guidance, it is imperative to take into consideration the self-agency of the users. In our post-study interview, participants stated that they would want some degree of control over the device facilitating spatial guidance such that they can cancel or modify the given instructions at any point. Moreover, we also believe that it would be critical to ensure that the afforded spatial guidance is gradual and suggestive, but not overpowering, and that sufficient movement information is provided to the user before initiating the spatial guidance, since sudden movements could affect the visually impaired user's ability to maintain their body balance [3, 36, 38].

9.3 Limitations and Future Work

While the idea of automated hand-based spatial guidance is promising in itself, the susceptibility of this technique to potential risks must not be overlooked. The technique takes the burden of interpreting the guidance off the user by directly controlling the user's hand. As the user builds trust with the guidance system, they may relax their agency entirely, which can be dangerous if the system gets compromised. While traditional feedback modalities are also susceptible to malicious agents, these guidance modalities are more suggestive in nature and grant the user agency in deciding whether or not to act upon the provided suggestions. Therefore, safety must be carefully considered when implementing automated hand-based spatial guidance. As the minimum measure of safety, the user should be able to override the automated hand actuation. In our study, users unanimously agreed that they felt completely in control over FingerRover, as they could stop following the spatial guidance at any point by simply lifting their fingers, and the device was not strong enough to override the user. However, this may not be the case with other implementation techniques such as EMS, and thus special attention needs to be directed towards agency-overriding mechanisms. Furthermore, users should be able to make an informed decision to stop or override the automated guidance. The study demonstrated that our method allows blind and low-vision participants to understand the movement of the robot and even interpret its trajectory. However, providing additional channels to communicate the status of the guidance, such as the current progress, may help users be more confident in using the system.

Our study also comes with some limitations. We conducted the study in a controlled lab environment with several assumptions such as perfect camera positioning and marker tracking system. Furthermore, we simulated the interaction design for each task based on the state-of-the-art solution available for that task, e.g., StateLens [20] for operating inaccessible touchscreen interfaces, since the goal of this study was to evaluate how directly manipulating the spatial modality could support interactions for the visually impaired. While our study showed promising results in a controlled setting, we believe testing these interactions in a more natural setting for a longer duration may reveal additional insights for adopting spatial guidance in real-world scenarios. Finally, we do not claim FingerRover to be the ideal form factor for facilitating automated hand-based spatial guidance. For instance, the current implementation of FingerRover can only move on a flat, horizontal surface. Future work could explore different form factors and actuation techniques for supporting automated hand-based spatial guidance in larger settings.

10 CONCLUSION

In this paper, we introduced *automated hand-based spatial guidance* – a technique that allows visually impaired users to move their hands between two points automatically, without the need to interpret audio/haptic directional cues. A user study conducted with blind and low-vision participants using *FingerRover* demonstrated the potential of using automated hand-based spatial guidance in helping the visually impaired perform daily tasks more efficiently and safely with less cognitive effort. The study also revealed various application scenarios where automated hand-based spatial guidance can be helpful. We hope our findings from this study can guide future research to further explore this modality of automated spatial guidance in improving assistive technologies for the visually impaired.

ACKNOWLEDGMENTS

We want to thank Anisha Gupta, Dana Draa, and Agatha Bisbikis for their help in recruiting participants for this study. We also want to extend our gratitude to all the participants for their involvement in this study and for their invaluable feedback.

REFERENCES

[1] Dragan Ahmetovic, Cole Gleason, Chengxiong Ruan, Kris Kitani, Hironobu Takagi, and Chieko Asakawa. 2016. NavCog: a navigational cognitive assistant for the blind. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 90–99. https://doi.org/10.1145/2935334.2935361

[2] Dragan Ahmetovic, Daisuke Sato, Uran Oh, Tatsuya Ishihara, Kris Kitani, and Chieko Asakawa. 2020. ReCog: Supporting Blind People in Recognizing Personal Objects. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376143

[3] Ahmad H. Alghadir, Abdullah Z. Alotaibi, and Zaheen A. Iqbal. 2019. Postural stability in people with visual impairment. *Brain and Behavior* 9, 11 (2019), e01436. https://doi.org/10.1002/brb3.1436 _eprint:

https://onlinelibrary.wiley.com/doi/pdf/10.1002/brb3.1436.
[4] Toshio Asano and Yutaka Ishibashi. 2007. A remote haptic painting lesson system. In *Multimedia Systems and Applications X*, Vol. 6777. SPIE, Boston, MA, USA, 295–302. https://doi.org/10.1117/12.746340
[5] Sandra Bardot, Marcos Serrano, and Christophe Jouffrais. 2016. From Tactile to Virtual: Using a Smartwatch to Improve Spatial Map Exploration for Visually Impaired Users. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 100–111. https://doi.org/10.1145/2935334.2935342
[6] Ulf Berglund and Birgitta Berglund. 1970. Adaptation and recovery in vibrotactile perception. *Perceptual and motor skills* 30, 3 (1970), 843–853.
[7] Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, and Tom Yeh. 2010. VizWiz: Nearly Real-Time Answers to Visual Questions. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology* (New York, New York, USA) *(UIST '10)*. Association for Computing Machinery, New York, NY, USA, 333–342. https://doi.org/10.1145/1866029.1866080
[8] Anke Brock, Slim Kammoun, Marc Macé, and Christophe Jouffrais. 2014. Using wrist vibrations to guide hand movement and whole body navigation. *i-com* 13, 3 (2014), 19–28.
[9] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh. 2021. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 43, 01 (jan 2021), 172–186. https://doi.org/10.1109/TPAMI.2019.2929257
[10] Elyse D. Z. Chase, Alexa Fay Siu, Abena Boadi-Agyemang, Gene S-H Kim, Eric J Gonzalez, and Sean Follmer. 2020. PantoGuide: A Haptic and Audio Guidance System To Support Tactile Graphics Exploration. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/3373625.3418023
[11] SeungA Chung, Kyungyeon Lee, and Uran Oh. 2020. Investigating three-dimensional directional guidance with nonvisual feedback for target pointing task. In *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, IEEE, Recife, Brazil, 206–210.
[12] Akansel Cosgun, E. Akin Sisbot, and Henrik I. Christensen. 2014. Evaluation of rotational and directional vibration patterns on a tactile belt for guiding visually impaired people. In *2014 IEEE Haptics Symposium (HAPTICS)*. IEEE, Houston, TX, USA, 367–370. https://doi.org/10.1109/HAPTICS.2014.6775483 ISSN: 2324-7355.
[13] Artem Dementyev, Hsin-Liu (Cindy) Kao, Inrak Choi, Deborah Ajilo, Maggie Xu, Joseph A. Paradiso, Chris Schmandt, and Sean Follmer. 2016. Rovables: Miniature On-Body Robots as Mobile Wearables. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 111–120. https://doi.org/10.1145/2984511.2984531
[14] Julie Ducasse, Marc Macé, Bernard Oriola, and Christophe Jouffrais. 2018. BotMap: Non-visual panning and zooming with an actuated tabletop tangible interface. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 4 (2018), 1–42.
[15] Tim Duente, Max Pfeiffer, and Michael Rohs. 2017. Zap++: A 20-Channel Electrical Muscle Stimulation System for Fine-Grained Wearable Force Feedback. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) *(MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, Article 1, 13 pages. https://doi.org/10.1145/3098279.3098546
[16] Shirin Feiz, Syed Masum Billah, Vikas Ashok, Roy Shilkrot, and IV Ramakrishnan. 2019. Towards Enabling Blind People to Independently Write on Printed Forms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300530
[17] Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Francisco José Madrid-Cuevas, and Manuel Jesús Marín-Jiménez. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47, 6 (June 2014), 2280–2292. https://doi.org/10.1016/j.patcog.2014.01.005
[18] Simon L. Gay, Edwige Pissaloux, Katerine Romeo, and Ngoc-Tan Truong. 2021. F2T: A Novel Force-Feedback Haptic Architecture Delivering 2D Data to Visually Impaired People. *IEEE Access* 9 (2021), 94901–94911. https://doi.org/10.1109/ACCESS.2021.3091441
[19] Anhong Guo, Xiang 'Anthony' Chen, Haoran Qi, Samuel White, Suman Ghosh, Chieko Asakawa, and Jeffrey P. Bigham. 2016. VizLens: A Robust and Interactive Screen Reader for Interfaces in the Real World. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. Association for Computing Machinery, New York, NY, USA, 651–664. https://doi.org/10.1145/2984511.2984518
[20] Anhong Guo, Junhan Kong, Michael Rivera, Frank F. Xu, and Jeffrey P. Bigham. 2019. StateLens: A Reverse Engineering Solution for Making Existing Dynamic Touchscreens Accessible. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 371–385. https://doi.org/10.1145/3332165.3347873
[21] Jonggi Hong, Alisha Pradhan, Jon E. Froehlich, and Leah Findlater. 2017. Evaluating Wrist-Based Haptic Feedback for Non-Visual Target Finding and Path Tracing on a 2D Surface. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '17)*. Association for Computing Machinery, New York, NY, USA, 210–219. https://doi.org/10.1145/3132525.3132538
[22] Jonggi Hong, Lee Stearns, Jon Froehlich, David Ross, and Leah Findlater. 2016. Evaluating Angular Accuracy of Wrist-based Haptic Directional Guidance for Hand Movement. In *Proceedings of the 42nd Graphics Interface Conference (GI '16)*. Canadian Human-Computer Communications Society, Waterloo, CAN, 195–200.
[23] Samantha Horvath, John Galeotti, Bing Wu, Roberta Klatzky, Mel Siegel, and George Stetten. 2014. FingerSight: Fingertip Haptic Sensing of the Visual Environment. *IEEE Journal of Translational Engineering in Health and Medicine* 2 (2014), 1–9. https://doi.org/10.1109/JTEHM.2014.2309343 Conference Name: IEEE Journal of Translational Engineering in Health and Medicine.
[24] Felix Huppert, Gerold Hoelzl, and Matthias Kranz. 2021. GuideCopter - A Precise Drone-Based Haptic Guidance Interface for Blind or Visually Impaired People. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 218, 14 pages. https://doi.org/10.1145/3411764.3445676
[25] Hyunki In, Brian Byunghyun Kang, MinKi Sin, and Kyu-Jin Cho. 2015. Exo-Glove: A Wearable Robot for the Hand with a Soft Tendon Routing System. *IEEE Robotics Automation Magazine* 22, 1 (March 2015), 97–105. https://doi.org/10.1109/MRA.2014.2362863 Conference Name: IEEE Robotics Automation Magazine.
[26] Yutaka Ishibashi and Toshio Asano. 2007. Media synchronization control with prediction in a remote haptic calligraphy system. In *Proceedings of the international conference on Advances in computer entertainment technology (ACE '07)*. Association for Computing Machinery, New York, NY, USA, 79–86. https://doi.org/10.1145/1255047.1255063
[27] Shaun K. Kane, Jeffrey P. Bigham, and Jacob O. Wobbrock. 2008. Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility (Assets '08)*. Association for Computing Machinery, New York, NY, USA, 73–80. https://doi.org/10.1145/1414471.1414487
[28] Roberta L. Klatzky, James R. Marston, Nicholas A. Giudice, Reginald G. Golledge, and Jack M. Loomis. 2006. Cognitive load of navigating without vision when guided by virtual sound versus spatial language. *Journal of Experimental Psychology: Applied* 12, 4 (2006), 223–232. https://doi.org/10.1037/1076-898X.12.4.223 Place: US Publisher: American Psychological Association.
[29] Masaki Kuribayashi, Seita Kayukawa, Hironobu Takagi, Chieko Asakawa, and Shigeo Morishima. 2021. LineChaser: A Smartphone-Based Navigation System for Blind People to Stand in Lines. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 33, 13 pages. https://doi.org/10.1145/3411764.3445451
[30] Guanhong Liu, Tianyu Yu, Chun Yu, Haiqing Xu, Shuchang Xu, Ciyuan Yang, Feng Wang, Haipeng Mi, and Yuanchun Shi. 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* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 28, 13 pages. https://doi.org/10.1145/3411764.3445644
[31] Pedro Lopes, Patrik Jonell, and Patrick Baudisch. 2015. Affordance++: Allowing Objects to Communicate Dynamic Use. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2515–2524. https://doi.org/10.1145/2702123.2702128
[32] Microsoft. 2010. Kinect for Windows - Windows apps. https://docs.microsoft.com/en-us/windows/apps/design/devices/kinect-for-windows
[33] Ken Nakagaki, Sean Follmer, and Hiroshi Ishii. 2015. LineFORM: Actuated Curve Interfaces for Display, Interaction, and Constraint. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 333–339. https://doi.org/10.1145/2807442.2807452
[34] Romain Nith, Shan-Yuan Teng, Pengyu Li, Yujie Tao, and Pedro Lopes. 2021. DextrEMS: Increasing Dexterity in Electrical Muscle Stimulation by Combining It with Brakes. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 414–430. https://doi.org/10.1145/3472749.3474759
[35] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. 2016. You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 779–788. https://doi.org/10.1109/CVPR.2016.91
[36] Izabela Rutkowska, Grzegorz Bednarczuk, Bartosz Molik, Natalia Morgulec-Adamowicz, Jolanta Marszałek, Kalina Kaźmierska-Kowalewska, and Krzysztof Koc. 2015. Balance Functional Assessment in People with Visual Impairment. *Journal of Human Kinetics* 48 (Jan. 2015), 99–109. https://doi.org/10.1515/hukin-2015-0096

[37] Oscar Sandoval-Gonzalez, Juan Jacinto-Villegas, Ignacio Herrera-Aguilar, Otniel Portillo-Rodiguez, Paolo Tripicchio, Miguel Hernandez-Ramos, Agustín Flores-Cuautle, and Carlo Avizzano. 2016. Design and Development of a Hand Exoskeleton Robot for Active and Passive Rehabilitation. *International Journal of Advanced Robotic Systems* 13, 2 (March 2016), 66. https://doi.org/10.5772/62404 Publisher: SAGE Publications.

[38] Micaela Schmid, Antonio Nardone, Alessandro Marco De Nunzio, Monica Schmid, and Marco Schieppati. 2007. Equilibrium during static and dynamic tasks in blind subjects: no evidence of cross-modal plasticity. *Brain* 130, 8 (Aug. 2007), 2097–2107. https://doi.org/10.1093/brain/awm157

[39] Meng-Li Shih, Yi-Chun Chen, Chia-Yu Tung, Cheng Sun, Ching-Ju Cheng, Liwei Chan, Srenivas Varadarajan, and Min Sun. 2018. Dlwv2: A deep learning-based wearable vision-system with vibrotactile-feedback for visually impaired people to reach objects. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, IEEE, Madrid, Spain, 1–9.

[40] Roy Shilkrot, Jochen Huber, Wong Meng Ee, Pattie Maes, and Suranga Chandima Nanayakkara. 2015. FingerReader: A Wearable Device to Explore Printed Text on the Go. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2363–2372. https://doi.org/10.1145/2702123.2702421

[41] Mathieu Simonnet, Anke M Brock, Antonio Serpa, Bernard Oriola, and Christophe Jouffrais. 2019. Comparing interaction techniques to help blind people explore maps on small tactile devices. *Multimodal Technologies and Interaction* 3, 2 (2019), 27.

[42] Patrick Slade, Arjun Tambe, and Mykel J. Kochenderfer. 2021. Multimodal sensing and intuitive steering assistance improve navigation and mobility for people with impaired vision. *Science Robotics* 6, 59 (2021), eabg6594. https://doi.org/10.1126/scirobotics.abg6594 arXiv:https://www.science.org/doi/pdf/10.1126/scirobotics.abg6594

[43] Sony. 2022. Toy platform toio™ / Stories / Sony Design / Sony. https://www.sony.com/en/SonyInfo/design/stories/toio/ Online; accessed Aug 2022.

[44] Lee Stearns, Ruofei Du, Uran Oh, Catherine Jou, Leah Findlater, David A. Ross, and Jon E. Froehlich. 2016. Evaluating Haptic and Auditory Directional Guidance to Assist Blind People in Reading Printed Text Using Finger-Mounted Cameras. *ACM Transactions on Accessible Computing* 9, 1 (Oct. 2016), 1:1–1:38. https://doi.org/10.1145/2914793

[45] Ryo Suzuki, Clement Zheng, Yasuaki Kakehi, Tom Yeh, Ellen Yi-Luen Do, Mark D. Gross, and Daniel Leithinger. 2019. ShapeBots: Shape-Changing Swarm Robots. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 493–505. https://doi.org/10.1145/3332165.3347911

[46] Saiganesh Swaminathan, Thijs Roumen, Robert Kovacs, David Stangl, Stefanie Mueller, and Patrick Baudisch. 2016. Linespace: A Sensemaking Platform for the Blind. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2175–2185. https://doi.org/10.1145/2858036.2858245

[47] Emi Tamaki, Takashi Miyaki, and Jun Rekimoto. 2011. PossessedHand: Techniques for Controlling Human Hands Using Electrical Muscles Stimuli. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 543–552. https://doi.org/10.1145/1978942.1979018

[48] Yudai Tanaka, Jun Nishida, and Pedro Lopes. 2022. Electrical Head Actuation: Enabling Interactive Systems to Directly Manipulate Head Orientation. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 262, 15 pages. https://doi.org/10.1145/3491102.3501910

[49] Sensable Technologies. 2003. "Phantom Omni" – 6 DOF master device | Delft Haptics Lab. https://delfthapticslab.nl/device/phantom-omni/ Online; accessed Aug 2022.

[50] Unity Technologies. 2018. Unity's AR Foundation Framework | Cross platform augmented reality development software | Unity. https://unity.com/unity/features/arfoundation

[51] Kaveri Thakoor, Nii Mante, Carey Zhang, Christian Siagian, James Weiland, Laurent Itti, and Gérard Medioni. 2015. A System for Assisting the Visually Impaired in Localization and Grasp of Desired Objects. In *Computer Vision - ECCV 2014 Workshops (Lecture Notes in Computer Science)*, Lourdes Agapito, Michael M. Bronstein, and Carsten Rother (Eds.). Springer International Publishing, Cham, 643–657. https://doi.org/10.1007/978-3-319-16199-0_45

[52] Nelson Daniel Troncoso Aldas, Sooyeon Lee, Chonghan Lee, Mary Beth Rosson, John M. Carroll, and Vijaykrishnan Narayanan. 2020. AIGuide: An Augmented Reality Hand Guidance Application for People with Visual Impairments. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3373625.3417028

[53] Ultraleap. 2013. Tracking | Leap Motion Controller | Ultraleap. https://www.ultraleap.com/product/leap-motion-controller/ Online; accessed Aug 2022.

[54] Bernhard Weber, Simon Schätzle, Thomas Hulin, Carsten Preusche, and Barbara Deml. 2011. Evaluation of a vibrotactile feedback device for spatial guidance. In *2011 IEEE World Haptics Conference*. IEEE, Istanbul, Turkey, 349–354. https://doi.org/10.1109/WHC.2011.5945511

[55] Zihan Wu, Chun Yu, Xuhai Xu, Tong Wei, Tianyuan Zou, Ruolin Wang, and Yuanchun Shi. 2021. LightWrite: Teach Handwriting to The Visually Impaired with A Smartphone. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 32, 15 pages. https://doi.org/10.1145/3411764.3445322

[56] Zhixuan Yu, Samantha J. Horvath, Alexandra Delazio, Jihang Wang, Rebeka Almasi, Roberta Klatzky, John Galeotti, and George D. Stetten. 2016. *PalmSight: an assistive technology helping the blind to locate and grasp objects*. Technical Report CMU-RI-TR-16-59. Carnegie Mellon University, Pittsburgh, PA.

[57] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. 2020. MediaPipe Hands: On-device Real-time Hand Tracking. arXiv:2006.10214 [cs.CV]

[58] Yuhang Zhao, Elizabeth Kupferstein, Hathaitorn Rojnirun, Leah Findlater, and Shiri Azenkot. 2020. The Effectiveness of Visual and Audio Wayfinding Guidance on Smartglasses for People with Low Vision. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376516

[59] Peter A Zientara, Sooyeon Lee, Gus H Smith, Rorry Brenner, Laurent Itti, Mary B Rosson, John M Carroll, Kevin M Irick, and Vijaykrishnan Narayanan. 2017. Third eye: A shopping assistant for the visually impaired. *Computer* 50, 2 (2017), 16–24.

A PRELIMINARY STUDY SURVEY RESULTS

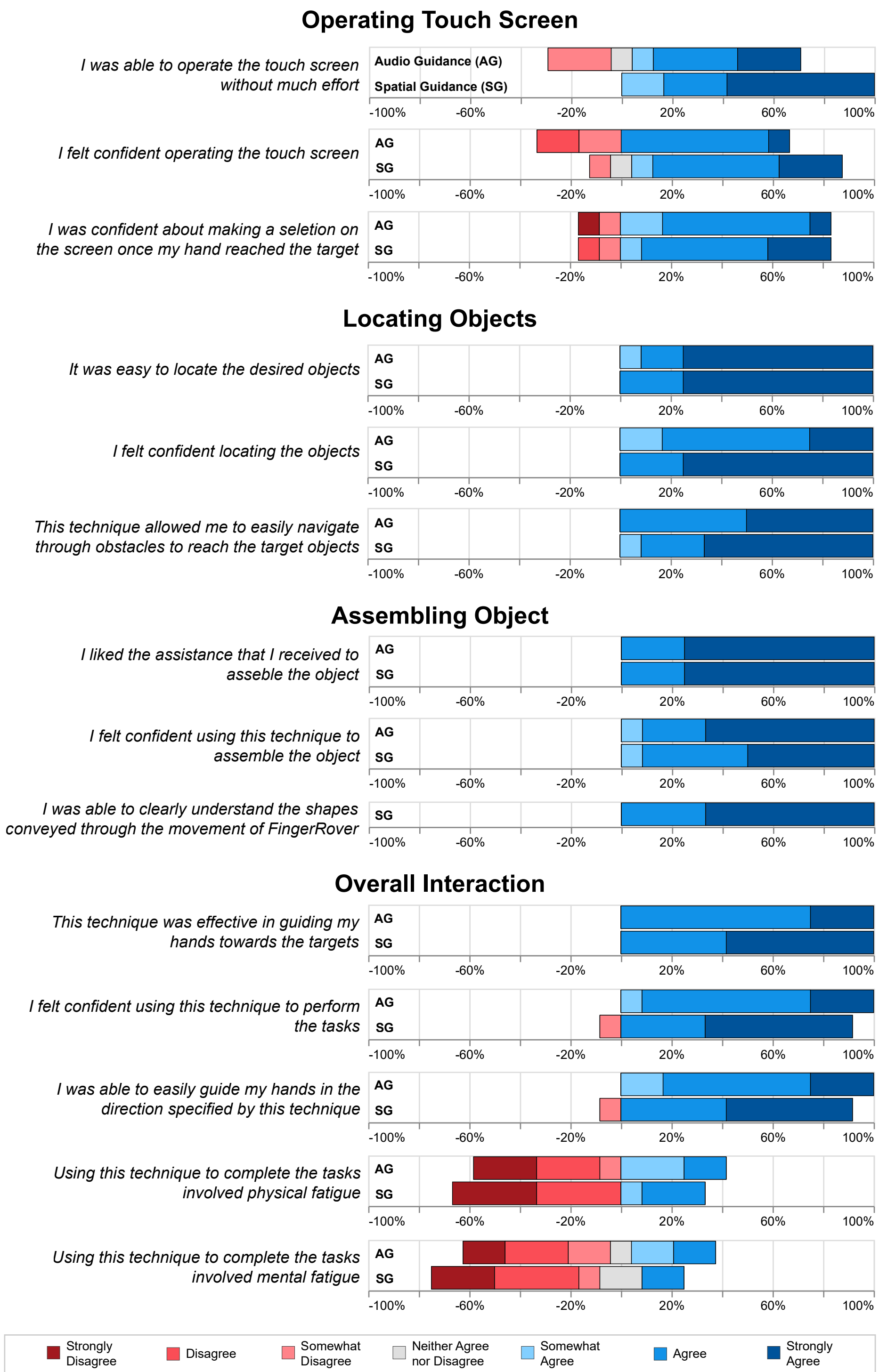

Figure 13: Blindfolded participants' Likert assessment for each interaction during the preliminary user study.