4-23-2020

The Emerging Professional Practice of Remote Sighted Assistance for People with Visual Impairments

Sooyeon Lee

Madison Reddie

Chun-Hua Tsai

John Millar Carroll

Mary Beth Rossen

See next page for additional authors

Authors
Sooyeon Lee, Madison Reddie, Chun-Hua Tsai, John Millar Carroll, Mary Beth Rossen, and Jordan Beck

The Emerging Professional Practice of Remote Sighted Assistance for People with Visual Impairments

Sooyeon Lee[1], Madison Reddie[1], Chun-Hua Tsai[1], Jordan Beck[2], Mary Beth Rosson[1], John M. Carroll[1]

[1]College of Information Sciences and Technology, The Penn State University, University Park, PA, USA

[2]Milwaukee School of Engineering, Milwaukee, WI, USA beckj@msoe.edu

ABSTRACT

People with visual impairments (PVI) must interact with a world they cannot see. Remote sighted assistance (RSA) has emerged as a conversational assistive technology. We inter- viewed RSA assistants ("agents") who provide assistance to PVI via a conversational prosthetic called Aira (https://aira.io/) to understand their professional practice. We identified four types of support provided: scene description, navigation, task performance, and social engagement. We discovered that RSA provides an opportunity for PVI to appropriate the system as a richer conversational/social support tool. We studied and identified patterns in how agents provide assistance and how they interact with PVI as well as the challenges and strategies associated with each context. We found that conversational interaction is highly context-dependent. We also discuss im- plications for design.

Author Keywords

Assistive technology; remote sighted assistance; visual impairment; human powered accessibility

CCS Concepts

- **Human-centered computing → Human computer inter- action (HCI);** *Accessibility;* Empirical studies in accessibility; Accessibility technologies;

INTRODUCTION

People with visual impairments (PVI) interact with their surroundings through various means, including their other sensory abilities, independent living skills, sighted people, and assistive technologies. A majority of assistive technologies (AT) and devices are for object recognition or mobility purposes. KNFB readers [1], barcode readers, Looktel money readers[2], color identifiers, Seeing AI[3], VoiceOver[4], TalkBack[5], talking thermometers, Microsoft Soundscape[6], BlindSquare[7], and Trekker Breeze are well-known examples. They compensate for lack of vision so that PVI can obtain the visual and directional information that they need.

However, most ATs convey visual information for only one specific context. For example, KNFB readers only read printed text and cannot read denominations of money. Soundscape provides only landmark information in outdoor settings, not turn-by-turn navigational information. Therefore, PVI often need to have and use multiple technologies to be accommodated in day-to-day life. Most of these ATs also do not adapt to changes, such as varying light, that can affect visual information's legibility. The limitations make these ATs non-ideal for PVI in dynamic environments. Moreover, visual impairments vary widely between visually impaired individuals. The complexity of user needs requires an advanced level of "intelligence," which most commonly available tools lack.

Most ATs cannot completely fulfill the wants and needs of PVI [10]. Humans, on the other hand, can comprehend and process contextual details and adapt to changing conditions as well as varying PVI preferences and needs. However, human assistance also has its disadvantages. PVI cite the following reasons for preferring not to get.

---

[1]https://knfbreader.com/

[2]http://www.looktel.com/moneyreader

[3]https://www.microsoft.com/en-us/ai/seeing-ai

[4]https://www.apple.com/accessibility/mac/vision/

[5]https://support.google.com/accessibility/android/answer/6283677?hl=en

[6]https://www.microsoft.com/en-us/research/product/soundscape/

[7]https://www.blindsquare.com/

assistance from sighted people: receiving human assistance from someone in the same physical space as them decreases the PVI's perceived level of independence, and PVI do not want to become a burden to family members and friends, which resonates with the findings of the study by Bigham et al. [17]. Additionally, sighted people are not always available, and PVI may not get to decide when and how long to receive sighted assistance.

For these reasons, the concept of remote sighted assistance (RSA) has received lots of attention in the PVI research com- munity. RSA avoids many of the disadvantages of purely technological AT and reliance on present sighted people or loved ones and simultaneously provides human intelligence-based assistance on-demand. It is seen as a hybrid AT that combines technology and human intelligence. Many prototypes from research and commercial efforts have been developed and evaluated, from the early idea of tele-assistance using information and telecommunication technology, to crowdsourced assistance using smartphone applications.

In this paper, we adopt a commercial RSA service, which is widely recognized by the PVI community, as our research platform. Our goal is to understand how and why their practice works, which will inform how future RSA can be improved. We contacted the company and were granted access to their professional RSA assistants ("agents"). We interviewed 14 agents, looking to answer the following research questions:

(1) What is the experience of providing RSA?
(2) What are the distinctive challenges agents face in key situations?
(3) What strategies do agents use in facing those challenges?

To our knowledge, we are the first research group to obtain such privileged access and the opportunity to learn about this practice from the professional agents' perspectives and re- port the findings. We identify patterns in how agents pro- vide assistance and interact with PVI. We further discuss how agents cope with challenges and use different strategies in four contexts, including *scene description*, *navigation*, *task performance*, and *social engagement*. The study revealed the complexity of PVI's needs and how RSA can help them to accomplish goals.

RELATED WORK

Remote Sighted Assistance Systems for PVI

Most research investigating RSA guidance for people with visual impairments has focused on navigation aids. To our knowledge, the earliest research to conceptualize the idea of RSA was the MoBIC Project with their mobility aid system for blind travelers [28]. They introduced the possibility of a digital map for RSA utilizing a geographical information system (GIS) and global positioning system (GPS), and the navigational information was delivered verbally.

Garaj et al. [14] implemented their RSA concept in a prototype and assessed its utility. Their prototype connected a live human assistant and PVI through video with a wearable digital video camera, GIS, GPS, and the assistant's personal computer. This system supported one-way communication of navigational information from the assistant to the PVI. An- other similar RSA system prototyped by Bujacz et al. [10] had the assistant provide navigation commands via video stream through a USB camera and two notebook computers with wireless internet connection. They studied the usefulness of the prototype in three cases - unguided, remotely guided, and non-remotely guided in a controlled, indoor setting. Their primary finding was that the system facilitated feelings of safety and enthusiasm among the blind participants. Branski et al. [2] evaluated the next version of this prototype with more advanced technology in a real-world setting. However, their study found that RSA was not useful in tasks such as crossing an intersection without traffic lights and during times of day with low light when video quality suffered.

The same researchers [2] refined the prototype further with the introduction of an initial form of bidirectional communication between the assistant and the blind user and more options of display arrangement on the RSA monitor. Their study emphasized the importance of training agents and recognized possible system acceptance issues with regard to portability and stigma from wearing the system visibly on the body. In addition to these lab-setting research studies, a number of commercial systems have been introduced, such as Visio.assistant by Video & Systems [10] and MicroLook by Design-Innovation-Integration [10].

Another RSA-based mobility aid for travelers with visual impairments was prototyped and studied by Scheggi et al. [31]. The distinction from previous examples is the use of vibro- tactile haptic feedback for providing directional information as opposed to verbal dictation. Chaudary et al. [12] also employed vibrational feedback in a video-based RSA system. It provides guidance through either vibrations indicating directions on a smart-cane or voice commands via a Bluetooth headset. Unlike the prior studies that focused on outdoor navigation, Rafian et al. [29] investigated the possible usefulness of an RSA system in indoor navigation using a crowdsourced indoor location sensing method. There is one study of RSA as a grocery shopping aid, TeleShop, by Kutiyanawala et al. [22].

All of these previous research efforts on RSA systems and the few commercial projects for PVI focused only on navigational problem solving. They did not fully facilitate two-way communication. A few studies explored the idea of two-way communication but in a very limited form, allowing one type of question and answer interaction. A research study by Holmes et al. [18] further explored two-way communication enabled by the video conferencing iPhone app, FaceTime, as an orientation tool for PVI.

Crowdsourced Remote Sighted Assistance Systems

As internet connectivity improved and the smartphone became more powerful, especially in camera functionality and quality, researchers started to investigate human crowdsourcing technology in RSA systems for PVI. One of the first research applications was the object recognition app, Vizwiz, developed by Bigham et al. [3]. It is a well-recognized smartphone app that accepts photos and questions from users and provides answers in text through crowd-sourced human assistance.

They have also used their initial findings to further develop and expand their work through VizWiz Social [17], which connected the user with visual impairment to their social networks, such as friends and family members, for assistance obtaining visual information. There are drawbacks of VizWiz, though, with many questions being deemed unanswerable by remote workers [3] due to the single photograph and question inputs, which are not suitable for complex and sequential contextual questions [24]. Researchers found that the blind VizWiz users want to partake in a synchronous interaction with the assistant as opposed to the text-based, asynchronous interaction VizWiz supported [3]. Furthermore, PVI are concerned about being a burden to their close acquaintances [24]. Brady et al.'s research [7] also examined social network-based, human-powered technology and presented the positives and negatives of the user experience and suggested design implications.

New research prototypes were developed by Lasecki et al. [23] to address the identified limitations. Legion [23] is the initial research project that introduced the concept of continuous, real-time interaction of crowd workers with the task through control of the user's interface. Building upon Legion, the re- search team developed a human, crowdsourced, conversational assistant, Chorus [25]. It employed an instant messenger to allow constant conversation and multiple crowd workers to collectively answer questions. However, usage was limited to web information-related inquiries, and the conversation is carried not verbally but in typed text messages.

The research team developed an improved version of Chorus, adding a video stream to facilitate a more continuous and conversational interaction and presented Chorus:View [24]. This prototype utilized FaceTime and a group of crowd workers who communicated with text. The researchers tested the prototype for tasks requiring multiple rounds of questions and answers. Even though Chorus:View addressed problems with prior prototypes, their study identified scalability and cost issues due to the video stream [24] and led to the development of another prototype, RegionSpeak [38]. It is a mobile application that uses image stitching and labeling for spatial layout understanding. It combines the best features of VizWiz and Chorus:View in that it provides both a simple photo-taking functionality and also context for multi-question interactions. However, they discovered that image stitching is difficult for cylindrical objects and latency could arise from the same labeling by a group of crowd workers.

The feasibility of crowdsourcing technology was also explored in a subjective information delivery context. Researchers used VizWiz to ask for subjective fashion advice from crowd workers [11]. User trust in non-expert volunteers to provide quality, subjective information was an issue. Another challenge was the availability of the crowd workers [11].

In addition to these academic research examples, there are commercial crowdsourced services existing as smartphone apps. TapTapSee [8] and BeMyEyes[9] are two such services that provide visual information to PVI. Although they have become well-recognized, the extent of the interaction between the crowd worker and the user and the scope of the assistance are very limited. They currently only perform object identification [6] and answer simple visual questions. The quality of answers and availability of crowd workers are not consistent nor guaranteed since it is a volunteer-based crowd.

Researchers analyzed current and historic examples of both academic and non-academic human-powered technologies and identified 13 design principles and considerations for designing human-powered systems [5, 6].

Collaborative Interaction between Sighted People and PVI RSA systems connect sighted assistants and visually impaired users, allowing them to interact collaboratively to achieve a shared goal. In general, collaboration is an important factor contributing to the success of a group activity. It is critical when a non-co-located sighted assistant and individual with visual impairment are performing a complex task. Examining how the two parties collaborate will help us understand the needs and challenges present in the RSA system space.

To our knowledge, there has been no in-depth research focused on this specific topic. However, there is work exploring the collaborative interaction between co-located sighted people and PVI. Williams et al. [35] studied outdoor navigation in blind-sighted pairs and reported how they approached navigation differently and what their findings imply for navigation AT design for PVI. Branham et al. [8] also explored collaborative accessibility created in shared home spaces between the blind and their sighted cohabitant. Yuan et al. [37] examined collaborative grocery shopping between the blind and different types of sighted partners and presented the necessary constituents of collaboration for successful shopping.

---

[8] https://taptapseeapp.com/

[9] https://www.bemyeyes.com/

Our literature review illustrates how RSA systems for PVI have been studied and improved and where the gaps lie. Most early research focused on navigation and object identification along with the advancement of camera technology and connectivity. Their findings were centered around the needs and challenges in RSA with regards to remote communication and the usefulness of their methods for specific tasks. Our study presents a more advanced RSA practice not only in the two areas extensively studied in the prior work but also in the space of task performance and social engagement, which is novel.

METHOD

Research Platform: Aira RSA System

In this study, we adopted *Aira RSA System* as our research platform. Aira[10] is a commercially available, on-demand RSA subscription service for PVI. As of 2019, Aira has offered service to thousands of "explorers" (users), advancing their learning, performance, and employment opportunities [11]. The service has been studied and associated with higher rates of year standing advancement of visually impaired students[12]. Aira's impact on the PVI community has been publicly recognized by several prestigious awards: TIME's Best Inventions of 2018[13]; Fast Company's 2019 World-Changing Ideas Award[14]; 2019 SXSW Interactive Innovation Award in the category of Social and Cultural Impact[15]; and two Mobile World Congress Global Mobile awards, the 2019 Best Innovation for Connected Life and Best Mobile Innovation for Education [16].

---

[10]https://aira.io/

[11]https://about.att.com/innovationblog/aira_school [12]https://healthcareweekly.com/aira-smart-glasses/

[13]https://time.com/collection/best-inventions-2018/5454219/aira/

[14]https://www.fastcompany.com/90329244/world-changing-ideas-2019-all-the-winners-finalists-and-honorable-mentions

[15]https://www.sxsw.com/awards/interactive-innovation-awards/#winners

[16]https://aira.io/mwc-2019

ID	Gender	Age	Time as an agent	Previous experience with PVI	Background/Occupation
Stella	F	25	10 months	None	N/A
Chris	M	50	1.5 months	Helped people work to regain lost vision in the past	Healthcare
Kevin	M	35	5 months	Has a friend who is PVI	N/A
Olivia	F	32	3 years	Grew up with a family friend who was blind	Criminal Justice
Jenn	F	23	1 year	None	Kinesiology
Brandon	M	27	6 months	None	Music
Ginger	F	22	10 months	Some volunteer work	Social Work
Kimberly	F	24	1 year	None	Aviation
Jessica	F	40	4 months	Grandfather was blind	Business Administration and Finance
Lily	F	27	4 months	Watched YouTubers who are part of PVI community	Retail, Customer Service
John	M	28	6 months	None	Musical theatre, TMC, EMS
Sarah	F	27	1 year	Knew someone who was blind growing up	N/A
Irin	F	23	1.5 years	None	Biochemistry
Ruby	F	23	5 months	None	Political Science, Freight Broker

Table 1. Overview of Participants (interviewees are anonymous due to privacy protection purpose.)

The professional agents are the primary reason we decided to adopt Aira as our research platform. The basic requirements of the RSA agents are (1) the agents are native English speakers; (2) the agents have general computer and internet literacy; and (3) the agents passed customer service aptitude tests and performed well in a simulated job environment involving customer service interactions, map reading, and working under time constraints. In addition to meeting the basic qualifications, all agents have been through two phases of training. The first is an online course describing company practices and culture, the blind community, orientation and mobility concepts and terms, and the technology used to administer the RSA. The trainees learn protocols and best practices like saying “you” rather than “we” in referring to ongoing activities, avoiding giving opinions, never speaking while PVI cross traffic intersections, and how to handle certain special cases (such as the request to help a PVI drive a car). The second phase of training involves hands-on practice in which the trainee provides guidance to an experienced trainer for navigation, shopping, using public transportation, and identifying objects.

Aira uses a mobile app or smart glasses to connect people who are blind or have low vision (PVI) with trained professional agents. The PVI and agent can establish two-way communication via real-time video chat. The agents provide assistance based on the context that PVI give them and the PVI’s profile. The profile contains information about the PVI’s visual impairment and guidance preferences. Aira’s agents are trained to assist PVI with various daily life tasks that require seeing, moving, doing, and interacting in personal spaces as well as social/public settings. Their guidelines are minimal, leaving most of the strategizing and problem-solving up to the individual agents. They exercise creativity and improvise to tailor assistance to each user and situation. This research plat- form provides us with the unique opportunity to observe the emerging professional practice of RSA for people with visual impairments.

Study Design and Recruitment

We recruited Aira RSA agents for interviews with the help of the company. We created a recruitment letter and an online Google recruiting form and sent these to Aira. The letter and form were distributed to all the agents (approximately 250) so that those who wanted to participate could sign up. The participation was voluntary with no compensation. 18 agents submitted the form to us, and we reached out to them directly for further communication. From these, we recruited a total of 14 study participants. Our interviewees are comprised of four males and ten females, and their ages range from 22 years to 50 years old. The participant table (Table 1) lists demographic and personal details. This study has been IRB-approved at the University.

Data Collection and Analysis

We conducted phone interviews with each of the 14 agents since they are geographically distributed. We first read a con- sent form to each interviewee and obtained verbal consent to proceed and audio-record the calls. The interviews lasted from 45 minutes to one hour and were audio-recorded in their entirety. We used open-ended, semi-structured questions as a guide for conducting the interviews. We asked the agents to share their experiences assisting PVI remotely, specifically with regards to types of assistance, types of interaction, challenges and strategies, and suggestions for system improvements. The following questions were asked to all interviewees: Could you compare and contrast the types of assistance you provide in different situations?; Could you tell me how your communication with explorers changes from situation to situation?; What are the challenges involved in providing RSA and why?; How do you deal with those challenges?; What could help you to assist explorers better? Our rich narrative data affirms that phone interviews can yield meaningful data.

The interviewer has been involved in research relationships with PVI since 2015 and is an active sighted member of the National Federation of the Blind (NFB). All other members of the research team are also sighted. The interviewer had tried the Aira application in a simple navigation scenario (walking through a door) before the interviews were conducted. Her experience with the PVI community was referred to through- out the interviews where appropriate to establish empathetic rapport. The interviewer had never previously met or spoken to any interviewees.

All 14 audio-recorded interviews have been transcribed. The data was analyzed with a bottom-up, iterative, thematic approach by multiple researchers [9]. The interviewer conducted the first round of coding analysis on the interview data. Based on the initial coding, using the interview questions as an anchor, high-level themes emerged and were grouped into more general categories. The different kinds of assistance, modes of interaction and communication, strategies and knowledge used, challenges, and the fact that all of these factors vary based on the situation at hand are common themes. Individual comments from the interviews were placed in only one category each. Two more researchers also reviewed and analyzed all of the raw data in parallel. The findings were then examined and fitted into the picture painted by the existing literature.

FINDINGS

From our study, we identified four distinctive contexts in which the Aira RSA agents provide assistance to PVI. Those con- texts are (1) scene description and object identification, (2) navigation, (3) task performance, and (4) social engagement. Information acquired and delivered may be visual, directional, or topical/how-to information. Assistance is provided through a collaborative interaction between the agent and the PVI. In this section, we present how the agents obtain and deliver information to PVI, cope with challenges, and establish and sustain effective interactions in each context.

Context 1: Scene Description and Object Identification For scene description/object identification, the most commonly requested task, agents are asked to provide visual information that is either objective or subjective. Delivering objective information most often consists of reading text in physical documents such as mail, menus, recipes, and textbooks; read- ing digital text; and describing spaces, scenery, images, or videos/movies. Subjective visual information may be opinions on outfits, descriptions of people (e.g., appearance, facial expression), and explanations of abstract artwork.

Information acquisition and delivery

For these tasks, getting a clear and precise view of the scene of interest is crucial. Agents often capture pictures from the video feed and zoom in to obtain the necessary visual information. To get a clean photo, agents work with PVI to adjust the camera to an optimal position. Once they have a satisfactory view of the visual information, they relay it to the PVI.

Important foundational knowledge

We found that agents need basic knowledge about the severity of the PVI's visual impairment, the onset of the PVI's visual impairment, the PVI's experience, and topical knowledge in a specialized area, such as science or math, if applicable, to successfully represent visual information to PVI in a meaningful way. We found that being aware of the nature of the PVI's visual impairment and their experience with various tasks or places is very helpful in the agent's determination of how best to communicate with the PVI. Our interviewees explained how they leverage these factors to communicate effectively and efficiently based on the PVI's understanding of the visual world. One agent interviewee, Stella, said:

"One of the things we ask... is how long they have been blind... If I see that someone has been blind for five years or ten years, I know I can talk about color or talk about different nuances. If I meet someone that is blind their whole life, I can use different descriptions like shape."

Challenges and strategies

Even though agents identified scene description as a relatively straightforward task, they still encounter several challenges. For instance, asking PVI to move their phone to adjust the video frame was called annoying and time-consuming. This maneuvering usually involves a number of back-and-forth adjustments, which can frustrate both parties. Although there is no technological solution to these difficulties, some agents creatively develop their own methods of simplifying the adjustments. Olivia shared her strategy with us:

"Kind of customizing your directions based on how they react to it... Even if it's supposed to be 6 inches, if I know they're moving a foot for every inch I say, instead of saying... 'Oh it's actually 6 inches,' because I know they're going to move it 6 feet, I'll just say, 'Can you move a half of an inch?' and then sometimes that makes it move 6 inches like I need it to."

Images and text can be obscured by a number of environmental factors even if they are in frame. Capturing still photos from the video feed and adjusting exposure using an Aira feature can help agents clarify some scenes. One agent developed another creative solution for relaying semi-obscured text. She told us that she reads the first and last bits of legible text to give PVI as much context as possible and determine if the rest of the text is relevant.

Agents are often asked to help with tasks that require or benefit from specialized knowledge. Describing something complex that they have little background information about is challenging. Kimberly explained this struggle:

"I have had a couple people call in for specific diagrams like, 'The teacher sent us this, we need to know it, but what am I looking at?' (PVI)... I wish I could have either had more experience in that area or been able to describe the material more clearly because it does take her a while to understand."

Another agent, Jenn, said:

"Especially if you're working on something that you don't have a lot of knowledge on, then you're going to be a little bit confused."

Agents have found that using analogies that are related to the PVI's experience is helpful. Ginger shared a story of a time she used this technique. We were impressed with how she discovered something she had in common with the PVI during their conversation and utilized it to describe complex visuals.

"When someone was taking... an advanced genetics class... I mentioned... 'This one looks like a sewing needle,' and she said, 'Oh I love sewing!' (PVI), and after that, I used sewing analogies like, 'This one looks like a pin cushion,' and, 'This one looks like a thimble,' just to try to make it a little bit more familiar to her."

When PVI ask what an assistant thinks, it requires judgement and an explicit opinion, which often makes agents uncomfortable. They are trained to be objective, but some questions are inherently subjective. A common question assistants receive is whether an outfit looks nice. The agents try not to convey whether they like it but attempt to make general recommendations based on common sense. One agent, Brandon, said that he would look for objective, external sources of judgements, such as a Google search, and share where his information came from with the PVI.

"A lot of times we get explorers that want to know if the outfits that they're wearing are good for a certain outing, like, 'Hey I'm going to a business meeting, does this shirt look good with these pants?' (PVI)... We try to stay very objective as well, you know,

'They're brown pants, this is a blue shirt, and it looks like you have a brown belt on with it.' They're like, 'Which shoes should I wear, black or brown?' I'd say, 'Well, generally you want to wear brown matching your brown belt, so brown shoes, brown belt...' We give them information and then they have the ability to decide for themselves from the information that's given."

Finally, providing descriptions of people can become awkward for agents when they are on speaker phone and the person they are describing can hear them.

"I feel uncomfortable if I don't know if I'm on speaker phone... because sometimes, you can give objective information about someone, and it can come off insulting. If I said, 'There's a large woman standing in front of you, so if you want to go around her, you have to go quite a bit to the left...' Obviously that could come off as pretty offensive."

Agents said that, in these situations, their strategies are being as objective as possible, avoiding being offensive while still being truthful to PVI, using specific descriptors for facial expressions such as 'slightly smiling' or 'toothy smile,' using synonymous words like 'prominent nose' instead of 'big nose' or 'heavy set' instead of 'big,' and using Google as an information source.

Interaction style

In scene description, where the main task is presentation of visual information, we found a more asynchronous interaction develops. Agents provide the visual information and PVI receive it. During the interaction, not much bidirectional communication happens. If there is some back-and-forth, it is mostly for the purpose of positioning the PVI's camera to properly frame the scene or clarifying what the PVI needs. The necessity for two-way communication is minimal compared to in the other contexts.

Context 2: Navigation

Navigation is identified by agents as the most challenging context. While navigating, PVI need information about their surroundings and obstacles as well as directions. The larger volume of information to be relayed and the dynamic nature make this task difficult for the agent. Common examples of navigation tasks are guiding PVI to a gate in an airport, navigating a campus, and finding a specific location in an unfamiliar area.

Information acquisition and delivery

To successfully navigate PVI, assistants need to know as much information as possible about the setting and route. Agents use Google Maps and the PVI's video feed to gather this information. Once the agents orient themselves to the location, they begin guidance, which includes obstacle notifications, scene description, and directions.

One interesting practice agents described is their "silence rule" (designed by Orientation and Mobility Trainers) for when PVI are crossing an intersection. The agents are not allowed to speak or provide any assistance while the PVI is crossing. The rationale behind this rule is avoiding interfering with PVI's Orientation and Mobility (O&M) training [27]. PVI are trained to rely on their auditory senses to avoid potential danger (e.g. listening for cars crossing an intersection). An agent's speech can divert some of the PVI's attention or overload their auditory sense and cause them to misjudge dangerous situations. Therefore, Aira's protocol is that agents must refrain from speaking while PVI cross the street. Brandon described this protocol:

"'I will remain silent as you cross...' This is the one time where agents do not say anything because it's imperative that the explorer can hear... everything that is happening in the inter- section... We're not a replacement for their orientation and mobility training... All of these things that they have been taught, to navigate crosswalks and things like that, that is connected to independence... We never want to take that away from them because then it takes away that independence."

Important foundational knowledge

PVI preferences and practices such as the use of a cane versus a guide dog and comfortable directional vocabulary (left/right, degrees, clock face directions) tell the agent what and how to communicate. Agents have access to these details through the PVI's profile displayed on their dashboard. They also report inferring urgency from implicit (tone, speed) and explicit cues from the PVI. Ruby said:

"I immediately try and figure out what is the level of emergence in this call. How fast do they need me to help them?"

Challenges and strategies

The biggest challenge for all of our agent interviewees, especially on public roads, is having to deliver too much information in real-time and keep pace with the changing environment and movement of the PVI. One agent, Kevin, expressed the stress he experiences when PVI move faster than he can de- scribe the environment:

“They are cranking, they’re walking faster than people who are sighted. I actually need them to slow down. Words take time to express meaning.”

To cope with this challenge, agents can adjust their pace and prioritize certain information. If the PVI is walking down a crowded street, obstacle-related information will be the agent’s focus. Agents may also ask the PVI to slow down or stop and look left and right so that they can reorient and recalibrate. These negotiations are generally successful.

Another challenging situation mentioned by agents is unexpected road conditions or construction that is not shown on the maps they are depending on. Coarse or poor maps of malls or buildings also limit the agent’s ability to give proper guidance. Agents are trained to respond to unexpected conditions with simulations but have no technological aids in these situations.

They must orient themselves based only on the video feed. In some public places with insufficient maps, agents utilize sighted people around the PVI. The agents sometimes find an employee and ask them for directions after agreeing on this approach with the PVI. Kimberly shared her experience navigating a PVI in an airport:

“Sometimes, in stores and in airports, just like with sighted people, you’d better ask the employees for directions... We would give the explorer an option, ‘Would you like to... look for an employee to ask? If not, I can keep helping you the best that I can.’ But sometimes, just like anyone, you just get so lost that it’s best to ask an employee.”

Interaction style

All the agents expressed that navigation is one of the tasks they feel is teamwork-based. Throughout the conversational interaction, the agent and the PVI collaborate to alter the PVI’s speed and the level of detail in descriptions through verbal and nonverbal communication. We found that these are adjusted as the interaction unfolds to optimize the experience. PVI letting the agent know what they know about the area or what they remember about specific landmarks is a great help. Trust was also identified as an important contributor to smooth and fast travel.

“I love when someone is confident! It makes my job so much easier. They’re more responsive. Sometimes when you tell a (PVI) [to turn], they’ll stop slowly and turn and they’ll keep walking and say, ‘Am I going the right way?’(PVI), whereas people who are confident will turn left. They’re just so responsive and confident in their skills and trusting in me.”

When the agents were asked when they felt the greatest degree of teamwork, most of them said while navigating with a guide dog user. The agent, the PVI, and the guide a dog form an effective team for navigating.

“The biggest team would be like... the teamwork between the Aira agent... and when they have a dog because the dog is able to do a lot of the obstacle avoidance, so all we need to do is description and general directions. And that is like a really solid team there.”

Context 3: Task Performance

In the context of task performance, a series of actions are taken to accomplish a goal. Examples agents shared with us are PVI doing a live presentation; teaching a class; cooking; building a treadmill; putting on lipstick; voting; making presentation slides, banners, and resumes; and fixing a broken printer, TV, or assistive device.

Information acquisition and delivery

As in the previous contexts, agents need to know the fundamentals of how to provide meaningful information at the proper level of detail to PVI. However, the need for topical knowledge also emerges. The role of collaboration increases because more elaborate and fine movements and actions need to be guided.

The agent firstly depends on their personal knowledge and skills. Their secondary resources are Google searches and Youtube. Irin explained why she uses Google:

“If it is a task that I don’t know how to do or have no experience in, first I will offer to Google how to help them... because ultimately the goal is to provide just a good experience to the explorer. We don’t want them to have to always call back because we don’t feel confident in a task... I’ll try and do it myself."

Irin implies that agents can help PVI with tasks they lack personal experience in by looking to outside resources.

Another agent, Jenn, added:

“There are actually a lot of manuals for different things that you can find by just doing a quick Google search... What’s required is just the ability to find the information."

Google searches and manuals can yield large volumes of in- formation, so agents must determine what and how to share with PVI. The amount of information, when to provide it, and how to communicate it is agreed upon by the agent and PVI, usually early on in the call. Clear and timely establishment of these factors primes these interactions for success.

Stella provided us with a detailed case. It was interesting to find that she and a PVI had set up a communication protocol and rehearsed before a presentation.

“Beforehand, I went through each of the slides with her, and then we talked about them. So, she said... ‘These are the key words or the key facts I want to know.’ So then when the presentation started, I gave her the words she needed. For example, it was about traffic... Right as a slide came on I said ‘Traffic and two pictures,’ and so she knew there were two pictures that came on and she knew to talk about it."

This unique protocol they established also involved pausing and cuing. During a presentation, explicit verbal communication is not an option for PVI. Therefore, they use other modes like pausing or cuing, which alerts the agent that they would like to be fed information. Stella reported that many PVI have developed this skill and learned to strategically and discreetly take pauses. Oftentimes, PVI employ them when they forget or miss something. Stella explained:

“They’re pretty good at knowing when they have missed some- thing. So then they’ll pause or cue me like, ‘And the second point on this is’ [‘is’ spoken slowly and stretched out] and then they’ll pause and I will say, ‘Popsicle’ as quickly as I can. Then they will continue and say, ‘It’s popsicles!’"

In other kinds of performance tasks, we found another interesting practice that agents use to build trust with PVI. They report that when they are Googling information or waiting for PVI to complete a step, they narrate what is happening and what they are doing as opposed to staying silent. Olivia said:

“We always let them know what we are doing. Part of being an agent is getting used to narrating what we’re doing."

Brandon described the importance of this effort:

“One of the main ways that we keep trust in the explorer is we keep them informed of what we’re doing... I would say, ‘Hang on a second, I’m just doing a quick Google search,’ or, ‘I’m pulling up the manual right now...’ Then as long as the explorer is informed, it helps to keep that trust because you’re not leaving them out to dry... it’s all about maintaining that confidence level but also that kind of openness. It’s a collaboration."

Challenges and strategies

Some tasks that PVI seek help with can be tedious, procedural, and time-consuming. Steps often build on each other, and it is important that each is carefully executed. Building a treadmill is an example with sequential steps. Lengthy processes and detailed movements can frustrate PVI. Helping PVI cope with these frustrations can be a challenge.

“If they’re also getting extremely frustrated and they’re not listening or responding to whatever I’m saying, it can also be really hard for me because I have to try to calm them down, but they don’t want to listen, so that can be really frustrating.”

Assistants do their best to remain as calm and neutral as possible as well as empathize with the PVI. They try to refocus them and bring their attention back to the task. Jenn described her problem-solving method:

“When working with someone that has gotten super frustrated, I would just... pause and just talk to them, be like, ‘Is this working?’ If they want more information from the specific things that I’ve pulled up, or on my end just try to look for another solution... just let them vent if it comes to that... being able to vocalize that, sometimes that’s just what people need. Just being able to give them that space, understand, and hear them out, and then bring them back to the task at hand. ‘Okay, it looks like we’ve been trying this solution, but it isn’t working. I’ll go ahead and look for something else if you don’t mind.’"

We heard from an agent, Irin, that patience plays an important role in handling frustration:

“The most important thing that the agent can do is just be patient because people can be really frustrated or do things in ways that, to someone who is sighted, doesn’t make a lot of sense. But it’s not up to us to decide if the way that they are doing something is correct. It’s up to us to to give them objective information.”

Another challenge is developing hand-eye coordination be- tween two people, with the agent’s eyes and the PVI’s hands. This is more difficult for small, precise movements, such as writing with a pencil and paper. Agents spoke about the difficulties of filling out forms and paper ballots because accurate navigation over small scales allows little room for error. John found it very hard to help a PVI fill out a ballot, especially getting the pencil lined up over the correct, small circle.

Performance tasks frequently require topical information. The agent may need to provide information that is outside the scope of common knowledge. This is challenging for agents but unavoidable to help PVI with a large range of activities.

We learned that one way agents cope with this is collaborating with other agents to help one another find solutions to problems in a Slack channel. They also trade calls that would benefit from domain knowledge if another agent has experience in the relevant area. Brandon provided details about how it works:

“What’s really cool is we... use Slack... so if we have a task that we have no idea how to do, which there are many of, then we can pretty much send it on the Slack channel and say, ‘Hey, I have an explorer who needs help with this, anybody know how to do this or has anybody dealt with this before?’"

Interaction style

Asking questions, asking for confirmation, and interpreting non-verbal cues, such as tone, all actively take place during these types of tasks. The interaction is centered around team- work. Agents let the PVI know what they themselves are doing and verbally reiterate what PVI are doing. This substitutes for visual feedback that sighted people constantly receive while performing an activity.

Context 4: Social Engagement

In social contexts, agents help PVI who are in public spaces or who are interacting with one or more other people. Agents discussed cases including a blind dad helping his young son do homework and write, a blind mom reading a bedtime story to her child, a married couple (visually impaired husband and sighted wife) taking a tour of a home, two friends playing guitar (one visually impaired and one sighted), a blind presenter interacting with an audience, and a visually impaired student in a class. The agent needs to consider not only the PVI but also the third parties.

Information acquisition and delivery

The agent's goal is to help the PVI interact with other people naturally and seamlessly. Agents need to quickly interpret the context of the interaction and adapt. Because they don't have the PVI's full attention, they may also need to communicate more creatively.

Seeking help for interacting with others makes this assistive technology unique. We found that the coordination between the agent and PVI is especially strategic, as the interaction has social implications for the PVI. Brandon shared details of his experience helping a dad assist his son with homework:

"The explorer and I came up with an agreement... 'What do you want me to say or what do you want me to help with? Because I don't want to take away from that connection between you and your son...' The roles kind of switched because it was up to me to help the son draw the certain letters, so... the explorer became a little bit more of the assistant, and I took over as primary helper because I could see what the son was writing. But it was a very awesome give and take... kind of being in a way like a ghost helper."

Challenges and strategies

Balancing between distracting and facilitating is challenging when the PVI is having a dialogue with another person or is around others. The agent usually only speaks in between the dialogue. Sarah explained:

"When people are in the classroom they do not like that [agent speaking aloud]... Sometimes they [PVI] don't even have me on a headphone set. Sometimes whenever I talk, I'm being broadcast... So that is a challenge, speaking as little as possible while also giving them an effective experience."

Interaction style

A unique interaction style seen in social settings is discreet conversation. For times like when a PVI is sitting in a class where verbal conversation is not appropriate, agents and PVI have discreet ways of communicating, using either a messaging feature provided by the company or hand gestures. In some cases, agents are asked to send text messages back, and in others, PVI are able to use headphones so agents can speak to them. Sarah shared her experience with text messaging PVI:

"Someone would send a message and let me know they were in a situation and they couldn't talk... so they would just send me a message to look up something online... and [I would] verbalize to them because they would still be able to hear."

There are a few standard hand gestures that all agents under- stand, signaling simple requests like "Read" or "Stop." Agents and PVI can establish more for individual calls if needed. Stephanie explained:

"There are different hand signals that the explorers can use without talking and tell us if they want us to describe things or read things... They can tell us what they want us to do without having to talk."

DISCUSSION AND DESIGN IMPLICATIONS

Through a thorough analysis of our data, we were able to identify several shortcomings and opportunities for improvement in this RSA service. In addition to other improvements, we recommend using AI/AT to supplement human assistants in several situations, which both alleviates challenges to agents and provides better assistance to PVI. We also further discuss some of the interesting nuances of these complex interactions and their implications.

Scene Description

While Aria agents can reference PVI profiles and ask PVI about their experience with a given task, they may need do- main knowledge in addition to this personal knowledge. One agent called this a major challenge for her. She was asked to read music for a PVI but did not have any experience doing so. Agents use search engines to learn on the fly or pass off calls to other agents if they are lucky enough to find one with the desired expertise, but these quick fixes do not guarantee success. One agent suggested an option for PVI to be able to schedule calls with specific agents who are best qualified to assist them. PVI would know that they were going to receive quality assistance, and agents would be faced with unfamiliar tasks less often.

Bigham et al.'s VizWiz projects [4, 3] and Jayant et al. [19] as well as our study of Aira found positioning cameras to obtain usable images for human assistants to be another challenge. The other research groups worked to mitigate the problem with technological approaches. These helped but did not entirely solve the problem. Aira, in contrast, utilizes continuous, bidirectional interaction, allowing for immediate feedback, error correction, and detailed and personalized adjustment guidance. While the agent and PVI can quickly converge on a usable image together and this is an advantage of this type of RSA, the process can be simplified with future research.

There is also room for improvement in reading tasks, especially for longer texts. Agents reported that reading for an extended period tires their voice, and humans cannot read aloud as quickly as some assistive devices. Many PVI train themselves to understand spoken language at very high speeds. This is a strength that they develop to compensate for their vision, and it should be utilized. This suggests a design opportunity for leveraging PVI's strengths, making agents' jobs easier, and making RSA more robust. There is already highly efficient reading AT that could be incorporated into RSA. Hu- man and AI/AT strengths are complementary, which indicates benefits of integration.

Navigation

We found that navigation is a particularly difficult task for agents. They can only say one thing at a time despite busy, dynamic settings. One way to decrease the work an agent is responsible for is to add an additional information provider who would take on some portion of the work. One agent brought up the possibility of two agents dividing work and assisting one PVI.

This is also an opportunity to utilize existing technologies, like having Google Maps provide directions while an agent describes scenery and obstacles and can step in in case of any unexpected occurrences. A concern with this approach is overloading the PVI's auditory channel and the two sources drowning each other out. A potential solution is the use of a different sensory channel to deliver some information. Utilizing the tactile sense to signal obstacles or directions is one feasible method. Extensive prior work [13, 26, 32] investigating the effectiveness of multimodal information presentation, supported by Wickens' multiple resource theory [34], attests to the potential benefit of adding a haptic channel to RSA. This is the direction of our future research.

Many of the same challenges we found in the navigation con- text have also been identified in prior research and development [15, 16, 20, 21, 30, 36]. Williams et al. [36] studied personal and contextual differences in PVI navigation practices and the resulting implications for navigation technology design. Existing navigation ATs have not been able to man- age the inherent dynamic and unpredictable nature of these tasks while also accommodating individual differences. How- ever, while still challenging, RSA agents were able to adapt to variance in these factors and meet the changing needs of PVI. Agents' utilization of PVI's other navigation tools (O&M skills, guide dog, cane) increases the efficiency and effective- ness of the assistance using already-present resources. This observation supports the design implication described in [36], emphasizing the value of PVI's O&M training. Further, we found that O&M training can be leveraged to improve the quality of RSA. This insight can inform the design of future navigation AT for PVI.

Task Performance

Tasks involving performance, such as presentations in front of live audiences, may require communication methods besides explicit speech on the PVI side. The communication protocol is established between the agent and PVI and may even be practiced before the actual task.

Fine-tuned coordination among the two parties can allow PVI to carry out a presentation just as a sighted person does. We found that success is not dependent on systematic knowledge but rather the agent's sharpness, discretion, and creativity, which indicates that not every agent can provide the same level of complex assistance. Inconsistencies can also arise from misunderstandings regarding signals and cues, and this can impair the performance. Performance assistance and quick interpretation of nonverbal cues could be incorporated into agent training to provide a more consistent resource to PVI. Improvement in this domain can open doors for PVI in task performance for which they may desire discreet remote assistance. Even with systematic training, there is always potential for misunderstandings and mistakes, and how to systematically resolve these problems in RSA interactions is a possible direction of future research.

Social Engagement

Agents have difficulty assessing social situations from re- motely and can feel uncomfortable. It can be challenging to anticipate when someone is going to speak if not everyone present is within the video frame, which leads agents to be more conservative and speak less. As a result, the quality of the RSA may suffer. PVI can compensate for this limitation by providing feedback to agents, perhaps through hand signals, to indicate when an agent is free to speak and when they should hold off. The existing hand signals for "Speak" and "Stop" are not currently utilized much in social settings but can be easily reappropriated for this purpose.

Agents could be further enlightened about how to best help PVI when they are engaged in social interactions if they are familiar with the third parties and have context for the PVI's interaction with them. This information could be accrued by agents as PVI have social interactions and added to the existing PVI profiles.

We were interested to find that some agents went beyond what they were requested to do and made an extra effort to help PVI "look good" in social scenarios. Without sight, it is easy to miss nonverbal tells that convey the mood of a social interaction, which can lead to missteps that may make a PVI look bad, feel awkward, or offend someone. The agents we interviewed displayed a great deal of empathy and investment in the success of the PVI they help, which undoubtedly leads them to provide the best assistance that they can. However, the agents may feel uneasy about telling a

PVI about disapproval from their peers, another gap that can be filled by AI. The studies by Anam et al. [1] and Tanveer et al. [33] demonstrated the potential of using AI for reading nonverbal cues and facial expressions for PVI.

The versatility of Aira's RSA system has led agents to be pulled into social scenarios and implicated into PVI's close relationships and intimate interactions (e.g. assisting a father to help his son do homework, helping a parent read to a child). The agents are cognizant of the fact that they are in delicate situations and that they have an affect on the relationships that they are inserted into. The social dynamics created by the use of RSA for intimate interactions poses interesting research questions for future study.

Collaborative AT

We found Aira's service to critically depend on personal knowledge (e.g. preferred level of detail) and domain knowledge, synchronous interaction, and teamwork. Generalization of findings from the example of Aira can be investigated in other collaborative AT applications, which operate on many of the same concepts. In particular, augmenting human assistive capacities with those of AI/AT and agents' methods for trust- building, problem-solving, and discreet communication can be studied in other ATs. Future work can also synthesize findings from PVI-specific ATs like Aira with work involving ATs for other disabilities.

Limitations

One limitation of our research is inherent sampling bias, as all of our interviewees were Aira employees who volunteered to participate. This group has unique expertise in RSA, which was crucial to this study. Even though the participants' identities were protected, their employment relationship with Aira may have affected how candid they were. However, our data shows a balance of positive and negative findings, and all agents seemed happy to discuss challenges and shortcomings in Aira's service.

CONCLUSION

From our interviews with the professional agents, we found four main contexts in which agents provide assistance to PVI: scene description and object identification, navigation, task performance, and social engagement. Each of these con- texts demands agents be consistent, creative, and adaptable to quickly make decisions and provide a satisfying experience to PVI. We found that the RSA system augments human capabilities by equipping remote assistants with advanced technologies. With the human element at the center of the RSA system, collaborative conversational interaction is supported and what RSA can offer to PVI is extended more than ever before. This provides opportunities to appropriate the RSA system beyond reading and way-finding into the space of performing complex tasks and engaging in social settings with the tools of sight and human cognition at their fingertips. From this study, we learned how professional agents practice RSA and overcome challenges, and we contribute design implications and insights for future RSA improvement.

ACKNOWLEDGEMENTS

We thank Aira for providing us access to their RSA agents and the agents for participating in the interview study. We thank Krish Gudarsani, Kathleen Byrd, Rachael Owen, and Yanting Wu for their assistance in data analysis. This paper is based upon work supported by the NSF Expeditions: Visual Cortex on Silicon CCF 1317560.

REFERENCES

[1] Asm Iftekhar Anam, Shahinur Alam, and Mohammed Yeasin. 2014. Expression: A dyadic conversation aid using Google Glass for people who are blind or visually impaired. In *6th International Conference on Mobile Computing, Applications and Services*. IEEE, 57–64.

[2] Przemyslaw Baranski and Pawel Strumillo. 2015. Field trials of a teleassistance system for the visually impaired. In *2015 8th International Conference on Human System Interaction (HSI)*. IEEE, 173–179.

[3] Jeffrey P Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, and others. 2010a. VizWiz: nearly real-time answers to visual questions. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. ACM,

333–342.

[4] Jeffrey P Bigham, Chandrika Jayant, Andrew Miller, Brandyn White, and Tom Yeh. 2010b. VizWiz:: LocateIt-enabling blind people to locate objects in their environment. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE, 65–72.

[5] Jeffrey P Bigham, Richard E Ladner, and Yevgen Borodin. 2011. The design of human-powered access technology. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 3–10.

[6] Erin Brady, Jeffrey P Bigham, and others. 2015. Crowdsourcing accessibility: Human-powered access technologies. *Foundations and Trends® in Human–Computer Interaction* 8, 4 (2015), 273–372.

[7] Erin L Brady, Yu Zhong, Meredith Ringel Morris, and Jeffrey P Bigham. 2013. Investigating the appropriateness of social network question asking as a resource for blind users. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 1225–1236.

[8] Stacy M Branham and Shaun K Kane. 2015. Collaborative accessibility: How blind and sighted companions co-create accessible home spaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2373–2382.

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[10] M Bujacz, P Baranski, M Moranski, P Strumillo, and A Materka. 2008. Remote guidance for the blind—A proposed teleassistance system and navigation trials. In *2008 Conference on Human System Interactions*. IEEE, 888–892.

[11] Michele A Burton, Erin Brady, Robin Brewer, Callie Neylan, Jeffrey P Bigham, and Amy Hurst. 2012. Crowdsourcing subjective fashion advice using VizWiz: challenges and opportunities. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 135–142.

[12] Babar Chaudary, Iikka Paajala, Eliud Keino, and Petri Pulli. 2017. Tele-guidance based navigation system for the visually impaired and blind persons. In *eHealth*

360°. Springer, 9–16.

[13] Adam Faeth and Chris Harding. 2014. Emergent effects in multimodal feedback from virtual buttons. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 1 (2014), 3.

[14] Vanja Garaj, Rommanee Jirawimut, Piotr Ptasinski, Franjo Cecelja, and Wamadeva Balachandran. 2003. A system for remote sighted guidance of visually impaired pedestrians. *British Journal of Visual Impairment* 21, 2 (2003), 55–63.

[15] Nicholas A Giudice and Gordon E Legge. 2008. Blind navigation and the role of technology. *The Engineering Handbook of Smart Technology for Aging, Disability, and Independence* 8 (2008), 479–500.

[16] João Guerreiro, Dragan Ahmetovic, Daisuke Sato, Kris Kitani, and Chieko Asakawa. 2019. Airport Accessibility and Navigation Assistance for People with Visual Impairments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 16.

[17] Danna Gurari, Qing Li, Abigale J Stangl, Anhong Guo, Chi Lin, Kristen Grauman, Jiebo Luo, and Jeffrey P Bigham. 2018. Vizwiz grand challenge: Answering visual questions from blind people. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3608–3617.

[18] Nicole Holmes and Kelly Prentice. 2015. iPhone video link facetime as an orientation tool: remote O&M for people with vision impairment. *International Journal of Orientation & Mobility* 7, 1 (2015), 60–68.

[19] Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P Bigham. 2011. Supporting blind photography. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 203–210.

[20] Hernisa Kacorri, Eshed Ohn-Bar, Kris M Kitani, and Chieko Asakawa. 2018. Environmental factors in indoor navigation based on real-world trajectories of blind users. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 56.

[21] Seita Kayukawa, Keita Higuchi, João Guerreiro, Shigeo Morishima, Yoichi Sato, Kris Kitani, and Chieko Asakawa. 2019. BBeep: A Sonic Collision Avoidance System for Blind Travellers and Nearby Pedestrians. In *Proceedings of the 2019 CHI*

Conference on Human Factors in Computing Systems. ACM, 52.

[22] Aliasgar Kutiyanawala, Vladimir Kulyukin, and John Nicholson. 2011. Teleassistance in accessible shopping for the blind. In *Proceedings on the International Conference on Internet Computing (ICOMP)*. The Steering Committee of The World Congress in Computer Science,Computer Engineering and Applied Computing (WorldComp), 1.

[23] Walter S Lasecki, Kyle I Murray, Samuel White, Robert C Miller, and Jeffrey P Bigham. 2011. Real-time crowd control of existing interfaces. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 23–32.

[24] Walter S Lasecki, Phyo Thiha, Yu Zhong, Erin Brady, and Jeffrey P Bigham. 2013a. Answering visual questions with conversational crowd assistants. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 18.

[25] Walter S Lasecki, Rachel Wesley, Jeffrey Nichols, Anand Kulkarni, James F Allen, and Jeffrey P Bigham. 2013b. Chorus: a crowd-powered conversational assistant. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. ACM, 151–162.

[26] Sooyeon Lee, Chien Wen Yuan, Benjamin V Hanrahan, Mary Beth Rosson, and John M Carroll. 2017. Reaching Out: Investigating Different Modalities to Help People with Visual Impairments Acquire Items. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 389–390.

[27] Carolina Martinez. 1998. Orientation and mobility training: The way to go. *SEE/HEAR. Texas School for the Blind. Consulta efectuada em* 29, 06 (1998), 2006.

[28] Helen Petrie, Valerie Johnson, Thomas Strothotte, Andreas Raab, Rainer Michel, Lars Reichert, and Axel Schalt. 1997. MoBIC: An aid to increase the independent mobility of blind travellers. *British Journal of Visual Impairment* 15, 2 (1997), 63–66.

[29] Paymon Rafian and Gordon E Legge. 2017. Remote Sighted Assistants for Indoor Location Sensing of Visually Impaired Pedestrians. *ACM Transactions on*

Applied Perception (TAP) 14, 3 (2017), 19.

[30] Daisuke Sato, Uran Oh, Kakuya Naito, Hironobu Takagi, Kris Kitani, and Chieko Asakawa. 2017. Navcog3: An evaluation of a smartphone-based blind indoor navigation assistant with semantic features in a large-scale environment. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 270–279.

[31] Stefano Scheggi, A Talarico, and Domenico Prattichizzo. 2014. A remote guidance system for blind and visually impaired people via vibrotactile haptic feedback. In *22nd Mediterranean Conference on Control and Automation*. IEEE, 20–23.

[32] André Schmidt, Mads Kleemann, Timothy Merritt, and Ted Selker. 2015. Tactile communication in extreme contexts: Exploring the design space through kiteboarding. In *IFIP Conference on Human-Computer Interaction*. Springer, 37–54.

[33] M Iftekhar Tanveer, ASM Anam, Mohammed Yeasin, and Majid Khan. 2013. Do you see what I see?: designing a sensory substitution device to access non-verbal modes of communication. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 10.

[34] Christopher D Wickens. 2002. Multiple resources and performance prediction. *Theoretical issues in ergonomics science* 3, 2 (2002), 159–177.

[35] Michele A Williams, Caroline Galbraith, Shaun K Kane, and Amy Hurst. 2014. Just let the cane hit it: how the blind and sighted see navigation differently. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*. ACM, 217–224.

[36] Michele A Williams, Amy Hurst, and Shaun K Kane. 2013. Pray before you step out: describing personal and situational blind navigation behaviors. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 28.

[37] Chien Wen Yuan, Benjamin V Hanrahan, Sooyeon Lee, Mary Beth Rosson, and John M Carroll. 2017. I Didn’T Know That You Knew I Knew: Collaborative Shopping Practices Between People with Visual Impairment and People with Vision. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 118.

[38] Yu Zhong, Walter S Lasecki, Erin Brady, and Jeffrey P Bigham. 2015. Regionspeak: Quick comprehensive spatial descriptions of complex images for blind users. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2353–2362.