

What do Blind and Low-Vision People Really Want from Assistive Smart Devices? Comparison of the Literature with a Focus Study

Bhanuka Gamage
bhanuka.gamage@monash.edu
Monash University
Melbourne, Australia

Thanh-Toan Do
toan.do@monash.edu
Monash University
Melbourne, Australia

Nicholas Seow Chiang Price
nicholas.price@monash.edu
Monash University
Melbourne, Australia

Arthur Lowery
arthur.lowery@monash.edu
Monash University
Melbourne, Australia

Kim Marriott
kim.marriott@monash.edu
Monash University
Melbourne, Australia

ABSTRACT

Over the last decade there has been considerable research into how artificial intelligence (AI), specifically computer vision, can assist people who are blind or have low-vision (BLV) to understand their environment. However, there has been almost no research into whether the tasks (object detection, image captioning, text recognition etc.) and devices (smartphones, smart-glasses etc.) investigated by researchers align with the needs and preferences of BLV people. We identified 646 studies published in the last two and a half years that have investigated such assistive AI techniques. We analysed these papers to determine the task, device and participation by BLV individuals. We then interviewed 24 BLV people and asked for their top five AI-based applications and to rank the applications found in the literature. We found only a weak positive correlation between BLV participants' perceived importance of tasks and researchers' focus and that participants prefer conversational agent interface and head-mounted devices.

CCS CONCEPTS

- **Human-centered computing** → **Accessibility technologies**;
- **General and reference** → **Surveys and overviews**.

KEYWORDS

smart devices, wearable, mobile application, augmented reality, virtual reality, recognition

ACM Reference Format:

Bhanuka Gamage, Thanh-Toan Do, Nicholas Seow Chiang Price, Arthur Lowery, and Kim Marriott. 2023. What do Blind and Low-Vision People Really Want from Assistive Smart Devices? Comparison of the Literature with a Focus Study. In *The 25th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '23)*, October 22–25, 2023, New York, NY, USA. ACM, New York, NY, USA, 21 pages. <https://doi.org/10.1145/3597638.3608955>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ASSETS '23, October 22–25, 2023, New York, NY, USA

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0220-4/23/10...\$15.00

<https://doi.org/10.1145/3597638.3608955>

1 INTRODUCTION

According to the latest World Health Organisation (WHO) statistics, 2.2 billion people have a near or distance vision impairment [95]. For sighted people, vision is the primary way they perceive their environment and the objects in it. Lack of information about their environment and the objects in it is a severe disabling consequence of blindness and low-vision (BLV) [30, 48]. As a consequence, a major focus of assistive technology research has been to provide BLV people with information about their environment so that they can travel independently and more confidently interact with the objects and people around them [31, 37, 43, 62].

Recent advances in deep learning have led to breakthroughs in artificial intelligence (AI), in particular in natural language processing (NLP) and computer vision (CV) applications [8, 94]. As a consequence there has been widespread interest in utilising computer vision based assistive technologies to provide BLV people with environmental information. These range from simple object recognition [18, 21, 52] to autonomous guidance systems [34, 68, 78] and have been employed on a variety of “smart” devices ranging from glasses [46, 51, 66] to robotic dogs [29, 49].

However, while on average more than 200 papers are now being published each year on this topic, it is unclear if this research actually addresses the real-world needs of BLV people. The overarching question we tackle is: **Does current research in such smart assistive technologies address the needs of the BLV users, and if not what needs should researchers address?**

We conducted a three step investigation to investigate this question. The first step was a scoping literature review [75] (Section 3). We identified 646 research papers published within the past two and a half years that investigated the use of computer vision based applications as an aid for BLV people to comprehend various aspects of their environment. We recorded the specific task and devices utilised in these studies, and the kind of interaction model. We also recorded the involvement of people who are blind or have low-vision (BLV) in the research and their role as prior research suggests that end-user involvement in the design of assistive technology leads to increased satisfaction and adoption by users [90].

In the second step of our study (Section 4), we conducted semi-structured interviews with 24 BLV participants. During these interviews, we identified the top five tasks that they would like a smart device to assist with, as well as the type of device they preferred. We also asked them to rank the usefulness of the tasks identified

in the literature. Lastly, we inquired about any concerns they had regarding the use of smart devices.

In the final step of our study (Section 5), we compared and contrasted the results of the literature review and the interviews conducted with BLV participants. To the best of our knowledge this is the first research comparing researcher focus with BLV participant priorities in the field of assistive smart devices. Our findings were as follows:

- There is only a weak positive correlation between BLV participants' perceived importance of tasks and researchers' focus on those tasks.
- The majority of studies (82%) did not involve BLV participants. However, for those studies that did we found a slightly higher correlation, in particular if BLV participants were involved during the research design stage.
- While all participants were excited by the potential of computer vision based smart assistive technologies, they identified a number of tasks such as filling in paper forms and empty seat detection that warrant more attention by researchers.
- Conversational agent type interfaces are preferred by participants, but there is limited research on this topic in the context of smart devices providing environmental information.
- Head-mounted devices are favoured by participants, but the choice of wearable device depends on the task and environment.
- There is a desire for universal devices or platforms that enable seamless usage without the need for device switching.

This study will provide valuable guidance for researchers working in the field of smart assistive technologies and enable them to focus on tasks and interactions that are prioritised by the BLV community. In turn this will make it more likely that the technologies developed will be of use and adopted by the community.

2 BACKGROUND

In this section we review prior research into the use of smart devices by BLV people. We define smart devices as personal assistive technologies that use AI-based computer vision to help an individual understand their surroundings. They might be smartphone applications, stand-alone devices (hand-held, canes) or wearables (belts, caps). We are particularly interested in identifying the tasks that these smart devices can perform to help BLV users. This can be as simple as detecting a wallet or as complex as guiding the user to buy their groceries.

Smart device technologies: Hundreds of papers have described computer vision based smart devices intended for use by BLV people. These include hand-held devices [15, 64, 89], smart glasses [46, 51, 66], smart caps [58, 84, 98], smart canes [16, 76, 96] and many more [17, 27, 54, 91]. These utilise image enhancement [38, 63, 104], question and answering [32, 97] and other forms of interaction to assist their users. Tasks include Face Detection [24, 60], Text Recognition [79, 82], Road Sign Detection [45, 57] and many more. Numerous commercially available AI smart devices have also emerged employing a variety of devices [4–6, 10–12, 35].

Several papers have reviewed these different devices and technologies [56, 61, 65, 80, 88, 93]. Li et al. [65] evaluated head-mounted displays (HMD) for assistive and therapeutic applications for BLV people. They reviewed 61 studies and classified the studies based on the type of HMDs, the visualisation approaches, the visual condition and if the evaluation was done using a user study. The main focus of their study was to analyse the current state of assistive HMDs and present emerging trends.

Patel and Parmar [80] reviewed 13 vision substitution devices and studies for BLV individuals that used computer vision and image enhancement. The authors noted that only a limited number of assistive devices were commercially available, and the majority of devices had not been tested under real-world conditions.

Khan and Khusro [56] conducted a review of smartphone-based assistive technologies for BLV people, emphasising the importance of technological advancements, inclusive interfaces, and collaborative efforts among medical specialists, computer professionals, usability experts, and domain users to maximise the potential of ICT-based interventions for the visually impaired.

Simões et al. [88] reviewed technologies and methods for assisting BLV people with indoor navigation. They classified these indoor positioning systems and discussed the advantages and disadvantages of each approach.

The review of Kuriakose et al. [61] focused on studies into multi-modal navigation for BLV people. They provide a comprehensive review of the different forms of devices and their modalities. In each of the studies, they also identified if the system was evaluated with BLV users.

In their systematic mapping review of 105 papers, Valipoor and Antonio [93] primarily centred their investigation on the object recognition, obstacle detection and depth detection aspects of assistive solutions. The review highlighted that recent advances in computer vision hold the potential to revolutionise the creation of advanced and high-quality assistive technologies for BLV users.

However, none of these surveys directly considered the needs of the BLV community.

User requirements: Only a small number of papers have explicitly examined the requirements of BLV users. In 2001, Duckett and Pratt [39] conducted a study to identify the higher-level research topics deemed important by BLV participants. Understanding the environment was one of the main topics identified. However, this study was conducted before the widespread adoption of smartphones and mobile devices and did not specifically consider AI-based smart devices.

Tapu et al. [92] investigated the perception of wearable assistive devices by the BLV community. They reviewed and categorised these devices into sensor based and camera based devices. The authors then conducted consultations with visually impaired individuals, researchers, and software developers to identify a set of features that a wearable assistive device must possess to gain acceptance within the BLV community. Processing speed, portability, robustness and friendliness were some of them. They then scored the devices in the review with these features and found that none of them satisfied all the criteria.

In 2020, Plikynas et al. [81] explored requirements for indoor navigation systems. They looked into 27 papers with different forms

Table 1: Keywords

Technology	Visual Impairment	Task
"augmented reality"	"sight loss"	description
"virtual reality"	blindness	caption
"mixed reality"	"low vision"	recognition
mobile	"vision loss"	"image enhance*"
wearable*	"vision aids"	
"smart*phone app*"	"impaired vision"	
	"visual loss"	
	"vision impair*"	
"smart glasses"	"visual* impair*"	
Google Scholar		
"smart phone application"	"vision impairment"	"image enhancement"
"smart phone app"	"vision impairments"	"image enhance"
	"vision impaired"	"image enhancing"
	"visual impairment"	"image enhancer"
	"visual impairments"	
	"visually impaired"	

Table 2: Inclusion & Exclusion criteria

Inclusion
Must utilise computer vision to understand the immediate environment of the user.
The research application should state it aims to aid the BLV community.
Exclusion
Devices that are not worn or carried by the user such as smart home devices. Applications where only ultra-sonic sensors are used as a mode of detection.
Pure VR experiences without any real world integration.
Simulations of different visual impairments using VR.

of hardware and sensors and proposed directions for future developments. During their interview stage, they asked blind participants about the five biggest problems when orientating/navigating indoors. Finding room numbers, finding elevators and reading numbers at the bank were the top three problems.

The most relevant study is that of Golubova et al. [44]. They identified tasks that BLV participants would like a smart sight aid to perform. They did by asking the participants to wear recording glasses that can be turned on to capture a moment where they would use a 'perfect sight aid'. Reading package labels, reading signs, reading print on TV and identifying medicine were some of the most reported situations. One insight from the study was that many participants prefer devices that can perform a spontaneous task over a device that they have to wear for long hours. A limitation of the study was that it implicitly focused on HMD devices and did not explore participant preferences for smart devices or ways of interacting with these devices.

To date we see that there has been no comparison between the tasks and smart devices that researchers have focused on and the needs of BLV people. Furthermore, apart from [44] there has not been a general investigation into the tasks that BLV people would like smart devices to perform, and no investigation into the type of smart devices they prefer and the kind of interaction they would like. These are the issues that we address in this paper.

3 STUDY 1 - SCOPING REVIEW

The first step in our investigation was a literature review to ascertain the various tasks and smart devices that researchers have focused on. As part of this review we also examined the extent to which BLV participants were involved in the research. Given our goal of exploring and mapping the broader topic of smart assistive devices based on technology, devices, interactions, and involvement of BLV individuals, we opted to conduct a scoping review. As outlined by Munn et al. [75], scoping reviews are particularly well-suited for examining how research is conducted on a specific

topic and identifying and analysing gaps in the knowledge base, and identifying the types of available evidence in a given field.

3.1 Literature Search and Review Methodology

We captured papers from four databases: Google Scholar, Scopus, Web of Science, and PubMed. Search was based on keywords capturing three criteria: Technology, Visual Impairment, and high level Task. The top table of Table 1 shows the keywords: keywords in the same column were combined with "OR" and the columns combined with an "AND" operator during the search process.

Our initial search was conducted in July 2022. We restricted our search to the past two and a half years, specifically from 2020 to July 15th, 2022. There were two reasons for this: First, the majority of AI-based smart device research has been carried out in recent years. Second, the sheer size of the literature—we initially looked at the last five years and found over 30,000 papers.

Initial search: Our initial search with the keywords from the upper table of Table 1 with Google Scholar, found 4,992 papers, and that with Scopus, Web of Science, and PubMed, resulted in 264 papers. After applying a title and abstract screening process based on the exclusion criteria outlined in Table 2, we narrowed down the results to 454 papers from Google Scholar and 194 papers from the other databases. Merging and removal of duplicates, resulted in 587 papers for full-text review. During this review process, we excluded additional papers that did not meet our exclusion criteria, resulting in a final set of 477 papers. Selection of papers were primarily carried out by a single researcher but in consultation with other members of the research team when unclear.

Followup search: After conducting the initial search we discovered a limitation in Google Scholar's ability to handle wildcard queries within quoted multi-word queries. Unlike other databases, Google Scholar appears to interpret the symbol "*" as a wildcard word within multi-word queries, rather than as a wildcard completion. This means that, for example, that a search for "image enhance*" will match "image enhance by" but only match "image enhancing"

inconsistently on Google Scholar, while this term will consistently match "image enhancement" on the other databases.

As a solution, we expanded each multi-word-wildcard query (bolded text in Table 1) into separate multi-word keywords, as shown in the Google Scholar section of Table 1 and repeated the Google Scholar search.

We identified 11,034 new papers that were not in the initial search. Since the second search was conducted after July 15, 2022, we also removed papers that were published after that date to ensure consistency between the two groups. Filtering by the date and applying the title and abstract screening resulted in 214 papers. After merging the new papers with the previous results, we found an additional 140 papers, bringing the total number of extracted papers from 477 to 617.

Furthermore, it came to our attention that the term "*smart glasses*" was inadvertently omitted as a specific keyword. To rectify this, a third search was conducted, resulting in the inclusion of 29 additional papers, bringing the total count from 617 to 646.

3.2 Data Extraction

The following information was recorded for each paper.

- **Task:** The low-level task(s) the research was designed to support.
- **Device:** Type of smart device used in the research.
- **Interaction model:** The high-level interaction model(s) from the user's perspective. We examined whether the device enhanced images for the user (Image Enhancement), identified and described the user's surroundings (Scene and Object Description), found or located objects requested by the user (Detection and Recognition of User-Requested Objects), or engaged in conversation with the user (Question Answering).
- **Involvement of BLV participants:** We determined whether BLV participants were involved or not. And, if they were, whether this was during the design phase, the evaluation phase, or both phases.

3.3 Results

Tasks. Low-level tasks were extracted from the research studies. Two researchers went through these and combined tasks if they were closely related. For example, a study was detecting fruits and vegetables and another detecting seafood, were combined into Food Recognition. It should be noted that some tasks encompass other more specific tasks. For instance, Drug Pill Detection and Indoor Sign Detection utilise Text Recognition, but were classified as distinct tasks because they involved additional classification steps to provide more comprehensive output to the users, rather than simply reading text.

After finishing the data extraction, two of the researchers categorised these low-level tasks into higher-level groups. The initial grouping was informed by the Australian Assistive Products for Persons with Disability Classification and Terminology [2], which is based on the WHO ISO Assistive Technology Standards and Guidelines [1]. We introduced a new category called "Assistive products for cultural and sports activities" as we felt that tasks in this category did not readily fit into the other categories. The low-level

tasks, their description and high-level groups are shown in Table 5 (in the Appendix).

Once the tasks were determined we counted the number of papers addressing each task. The counts are summarised in Figure 1. It is worth noting that some papers addressed multiple tasks, such as Currency Recognition and Face Detection. Therefore, the total number of papers for all high-level categories adds up to more than 646 papers.

Our count revealed that the category of assistive products for handling objects and devices had the highest number of papers, accounting for 42.7% of the studies, followed by assistive products for personal mobility (40.4%), and communication and information (32.2%). The Object Detection & Localization task had the highest number of papers. The majority of these studies we reviewed were trained on well-known datasets such as COCO [67] and ImageNet [36], which include hundreds of object classes. As a result, these devices detect and share all objects from their training dataset, regardless of their relevance to the user.

Device. We collated a list of smart devices used in the studies and counted the number of papers that used each device. In cases where smart devices ran on Raspberry Pi or similar computing devices but did not specify the wearable location, we categorised it as a body-mounted device. Additionally, we classified smart-glasses and devices with head-mounted cameras as head-mounted devices. It is worth noting that some studies may fall under more than one category, such as a head-mounted device that utilises a smartphone app for voice input. In such cases, the study is counted for both wearables.

Figure 2 depicts the number of studies conducted for each type of wearable device. We observed 24 distinct types of devices in our review, with only 9 of them having at least 10 studies. The majority of studies employed smartphones as the primary device. This can be attributed to the fact that smartphones have all three essential components (i.e., camera, microphone, and computing unit) integrated into a single device as identified by Plikynas et al. [81] as well. Additionally, the widespread adoption of smartphone software development kits (SDKs) may have played a role in their popularity compared to more specialised tools required for other wearables.

Interaction model. The majority of studies (85%) focused on Scene and Object Description, while Question and Answering received the least attention (as shown in Figure 3). This could be attributed to two reasons. Firstly, wearable devices have limited computing power, which poses a challenge for developing on-device Question and Answering systems. Secondly, many studies were focused on the quantitative performance of the computer vision algorithms rather than user interaction.

Involvement of BLV participants. After collating the results, our analysis of the reviewed papers revealed that the BLV participants were not involved in the majority (82%) of the studies and only 105 out of the 646 reviewed had BLV participants—see Figure 4. We further examined the involvement of BLV participants at different stages of the studies. We found that in the majority of papers, BLV involvement was restricted to evaluation and only 38 papers out of

Assistive products for handling objects and devices (276)	General Objects (236)	Object Detection and Localization (231)
		Object Manipulation Aid (9)
	Specific Objects (47)	Currency Detection (25)
		Personal Object Recognition (13)
Others (16)		
Assistive products for personal mobility (261)	General Navigation (140)	Obstacle Detection (101)
		Safe Path Detection (11)
		Others (22)
	Outdoor Navigation (71)	Zebra Crossing Identification (12)
		Pedestrian Detection (11)
		Others (47)
	Indoor Navigation (44)	Sign Detection (12)
Room Detection (2)		
Shop Isle Guidance (2)		
Assistive products for communication and information (208)	Textual and Graphical Information (95)	Text Recognition (89)
		Barcode & QR Code Reader (4)
		Others (10)
	Human Interaction (79)	Face Detection (51)
		Emotion Detection (9)
		Others (26)
	Environmental Information (64)	Virtual Assistant (19)
Image Captioning (15)		
Others (24)		
Assistive products for personal care and protection (14)	Fall Detection (6)	
	Fire Detection (2)	
	Thermal Hazard Detection (1)	
Assistive products for cultural and sports activities (14)	Artwork Accessibility (6)	
	Museum Accessibility (4)	
	Physical Games Accessibility (3)	
Assistive products for personal medical treatment (4)	Drug Pill Detection (4)	

Figure 1: Breakdown of the counts of papers identified for each types of tasks. Note the height of each sub-box is only indicative of the number of papers in that category.

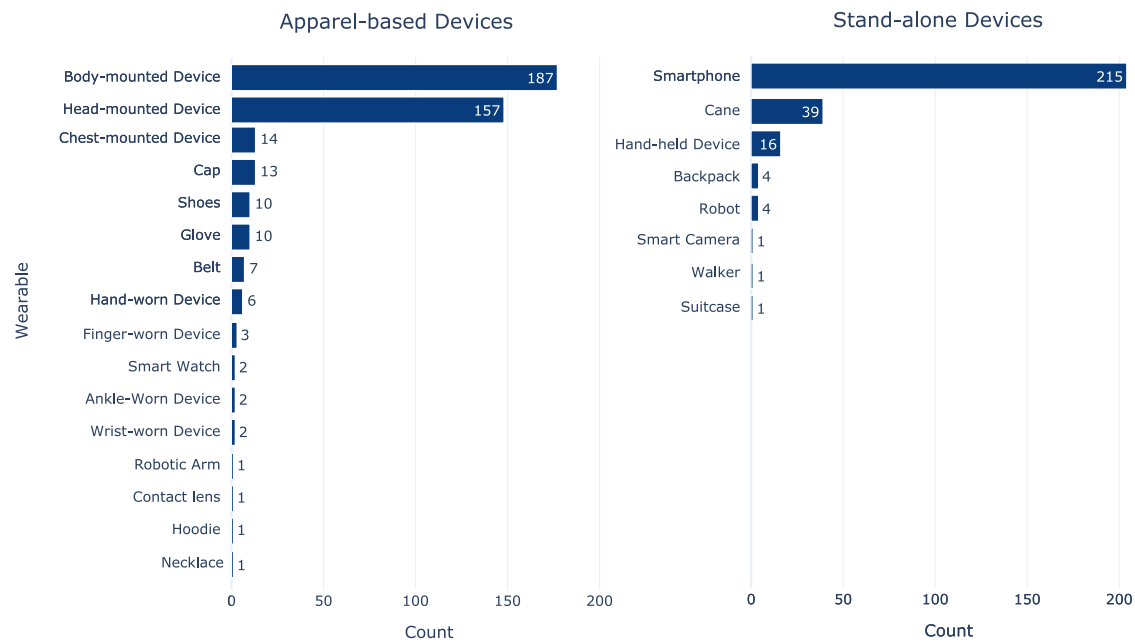


Figure 2: Count of papers identified for the types of devices

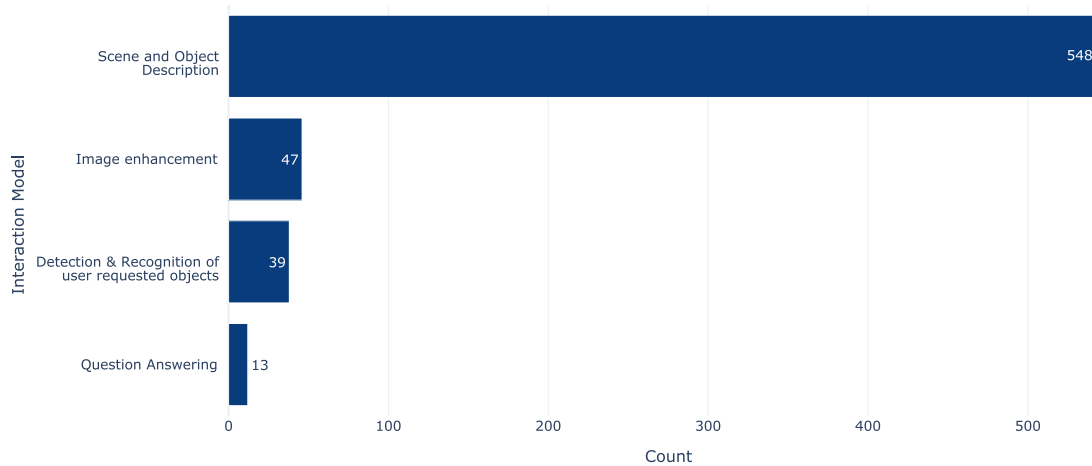


Figure 3: Count of papers identified for each interaction model

the 646 reviewed involved BLV participants in the design, ideation, or requirement gathering stage.

4 STUDY 2 - NEEDS OF PEOPLE WHO ARE BLIND OR HAVE LOW-VISION (BLV)

The second step in our research was to conduct semi-structured interviews with BLV participants to ascertain what they thought were the most useful tasks that smart devices could be used for. As part of this we asked about preferred interaction models and devices as well as asking them to rank the importance of the tasks that had been studied by researchers.

4.1 Participants

We conducted semi-structured interviews with 24 participants who either have low-vision or are blind. These lasted approximately one hour and participants were compensated with a \$50AUD gift card for their time. Of the 24 participants, 10 had low-vision and 14 were blind, with varying characteristics across age groups and age of onset. Table 4 (in the Appendix) summarises the demographic details of the participants, including age, level of vision, age of onset, and level of tech adoption.

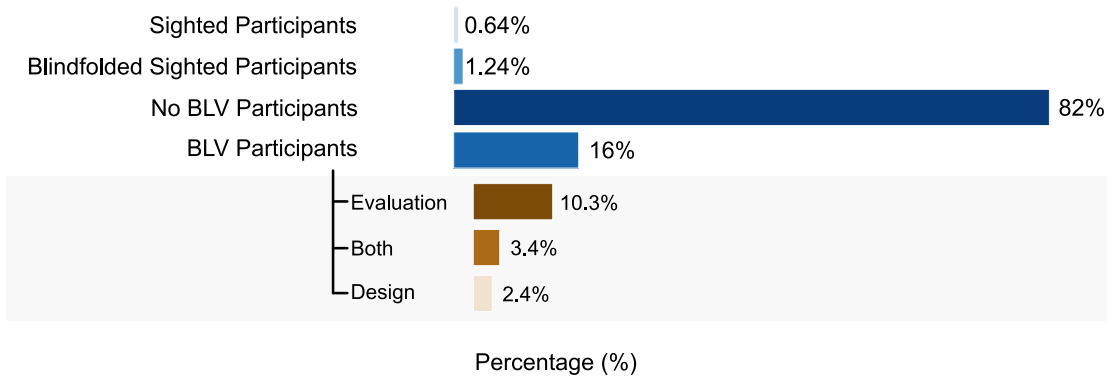


Figure 4: Type of participant involvement in research studies and the stages of involvement

4.2 Procedure

The interview had 5 parts.

Demographic information: At the beginning of each interview, a series of general questions were asked to ascertain the participants' demographic information and visual condition. Additionally, participants were requested to rank their level of technology adoption, which was categorised according to the summary provided by [22] as shown in Table 3 (in the Appendix).

Current technology use: Next, we asked about the types of devices participants currently utilise in their daily routines. If a participant mentioned using a smartphone or tablet, we further inquired about the applications they used on these devices.

Then we asked participants whether they have utilised any services that provide sighted assistance through video conferencing, such as AIRA [13] and Be My Eyes [3]. For those who have used such services, we followed up with questions to identify the tasks for which they utilised these services. Our objective was to identify tasks for which BLV individuals currently require sighted assistance, with the goal of determining if smart devices could potentially assist with these tasks. Lastly, we inquired whether any of the participants had any concerns regarding these services.

Most useful tasks: Prior to asking participants to list the tasks they felt would be the most useful, we provided them with a brief description of the types of smart devices, including smartphones, head-mounted devices, and chest-mounted devices. Additionally, we explained how artificial intelligence and machine learning has advanced in terms of mimicking human intelligence and how it can be incorporated into these devices. This was done to ensure that the participants had a general understanding of what constitutes a smart device. We carefully refrained from sharing any details that we identified from the review about the different tasks that these devices could perform, as we first sought to obtain the participants' perspective before presenting our findings from the review.

Next, we asked participants to list their top five tasks that they would like to accomplish using a smart device, followed by their suggestion for the type of smart device that would be best suited to the task.

In order to understand the two forms of modalities, i.e., human-computer (input) and computer-human (output) [73], we asked participants how they would prefer to interact with the device, including input methods such as voice, tactile buttons, and touch screen, and how they would prefer the device to provide information to them through output methods such as audio, visual or vibrations.

Ranking of researcher tasks: After this, participants were asked to rank the various tasks that researchers had focused on, as described in Table 5. To keep the interview to around one hour we could not ask each participant about all of the tasks. Instead we divided the participants into four groups, with each group ranking 18 tasks. We selectively assigned participants to each group to balance out the representation of low-vision and blind participants, ensuring that each group had at least two low-vision participants. To prevent any potential biases that may arise from the order in which the 18 tasks were presented to the participants during the interview, we randomised the task order for each individual.

While reviewing each task, the participants were asked to rank its importance based on its usefulness, ranging from "not useful", "slightly useful", "very useful", to "invaluable." Following this, they were asked a few questions to understand why the task was significant as the reason might differ from the researchers' original intention. Finally, they were requested to choose the 'best' type of smart device that would fulfil the task.

After the completion of the task ranking process, we provided the participants with an opportunity to re-evaluate their initial top five tasks to ensure that they had the chance to adjust their priorities based on the insights they gained from the ranking process.

Overall perception of smart devices: Following this, we asked the participants to rank the overall usefulness of smart assistive devices on the same scale as before, in order to gain a better understanding of their general perception of these devices. Lastly, we followed up with a question to discuss any concerns they may have regarding the development of smart devices for BLV individuals.

4.3 Results

Current technology use: All participants currently use a smart phone. This is consistent with Plikynas et al. [81], where they identified smartphones to be the most common electronic travelling

aid. OrCam (3), Envision glasses (2), BuzzClip (1), StrapTech (1), MiniGuide (1) and Trekker Breeze (1) were some of the other devices used. Note that the number in () is the number of participants using the device. We also found that Seeing AI (14), Google/Apple Maps (13), Soundscape (8), Be My Eyes (7) and Envision (6) were the most commonly used smartphone applications.

Out of the 16 participants who have used video conferencing services for sighted assistance, we identified the most frequently requested tasks. Reading text was the most commonly requested task by both blind and low-vision participants with 10 participants requesting it. Other frequently requested tasks included colour detection, website accessibility, navigating new environments, cleaning broken glass, framing and taking photographs, finding personal items, and assessing their physical presentation.

The results also revealed that our blind participants use these services more than low-vision participants (11 out of 14 blind participants use these services compared to 5 out of 10 LV participants). Out of the participants who used these services the blind participants tended to favour using AIRA over Be My Eyes, while low-vision participants preferred Be My Eyes over AIRA. This could be explained by the fact that low-vision individuals may rely on their limited vision to assist them with tasks, and the crowd-sourced support workers used by Be My Eyes may be sufficient for their needs. In contrast, blind individuals may require more specialised assistance and may therefore prefer the professionally trained staff provided by AIRA.

Finally, the majority of participants expressed positive feedback regarding the sighted assistance services, appreciating their convenience and professionalism, particularly when they lacked physical assistance. Only two participants raised concerns regarding issues such as privacy, connectivity, and high costs.

Most useful tasks: We compiled the responses of participants regarding the top 3-5 most preferred tasks and mapped them to the lowest level task classification shown in Table 5. For instance, one participant mentioned "something that can identify people" which was then mapped to Face Detection.

The most preferred task was Text Recognition, which was indicated by 15 out of the 20 participants, as demonstrated in Figure 5. Many participants reported that when they are outdoors, they often have to rely on others to read text for them, indicating the need for a reliable smart device to perform this task. The results of our study align with the findings of Golubova et al. [44], who reported that 18 out of their 32 participants identified reading any type of text as the most commonly named activity of importance. This finding also aligns with the results described in Section 4.3, where reading text was found to be the most requested task from video conferencing services.

We also identified that Obstacle Detection had the highest number of first preferences, with six participants selecting it. One participant, P9, reported that Obstacle Detection while navigating was the most challenging and stressful task for them. Therefore, a smart device that could aid in this area would greatly enhance their independence.

Figure 5 highlights several tasks in bold, including Empty Seat Detection, Sorting Items and Checking Appearance, that were not

included in the list of tasks from the review. We also included Navigation and Guidance as a separate task though it does overlap with Safe Path Navigation. This was primarily due to participants mentioning the need for Night Time Navigation and Guidance, which was not fully captured under Safe Path Detection. It is also because Navigation and Guidance is a higher-level task that combines several low-level tasks, such as Shop Recognition, Building Door Detection, and Safe Path Detection, to provide seamless interaction with the user, as demonstrated by [55].

We could not identify differences in the choice of top five tasks between the blind and low-vision participants.

Participant's preferred interaction mode: Figure 6 shows the different input and output modalities identified by the participants. It gives the number of participants that stated this modality for at least one of their tasks. In terms of preferred modes of interaction, 23 out of the 24 participants expressed a preference for voice input for at least one of their top tasks. This finding is also reflected in [50] where they compare the use of voice input by blind participants over sighted participants and found that blind individuals use voice input more frequently. [20] also found that 6 out of 8 blind participants preferred voice input over on-screen keyboard on mobile devices.

Participants also noted that, in the case of a smart glass device (head-mounted), voice input would be significantly more convenient for issuing queries, whereas a handheld device might be better suited to button-based input. Other popular input modalities included, 'user only needs to localise' and 'context-triggered automatic input' and button input. P4 described their ideal face-detection smart glass as having the capability to automatically track individuals they interacted with, capturing their names and periodically reminding the user when a known or unknown person was addressing them.

All participants indicated a preference for speech output as the primary means of interaction for at least one of their top tasks. Four participants expressed a preference for non-speech modalities such as beeps and tones, as well as vibration, due to concerns that speech output might interfere with their reliance on surrounding environment sounds. This finding aligns with the findings of [81] who found that participants did not want a device that would block their hearing when navigating indoor. However a factor impacting on the use of speech output is the need to avoid overwhelming the user. In particular, one participant (P14) emphasised the importance of the device being clear and concise when communicating with the user, as an overly verbose device could be overwhelming and interfere with task performance.

Comparing between the blind and low-vision groups, for input modalities, 7 blind participants showed a preference for the 'user only needing to localise', while only 3 low-vision participants exhibited this preference. In terms of output modalities, 3 low-vision participants preferred the visual modality, whereas (as one might expect) none of the blind participants did so.

Participant's preferred device: As depicted in column B in Figure 10, the top three wearable types preferred by the participants are head-mounted devices, smartphones, and hand-held devices. It is important to note that during the study, participants utilised the terms "hand-held devices" and "smartphones" interchangeably, where as certain participants explicitly referred smartphone applications.

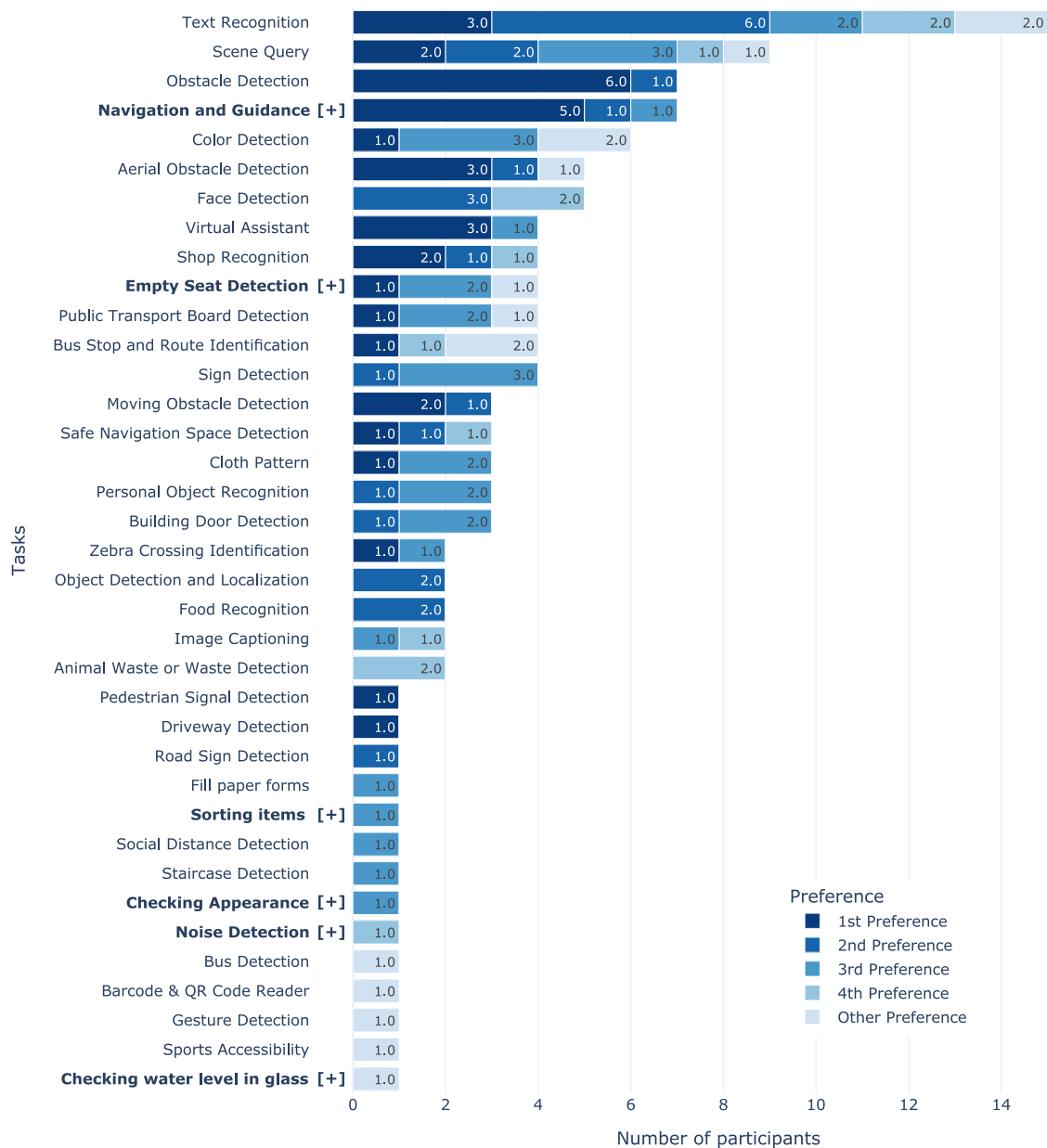


Figure 5: Participants' Five Most Useful Tasks Ranked by Order of Preference; The identified tasks are presented in order from most to least preferred, with tasks not listed in our review task list indicated in bold font with a [+] symbol next to them.

The participants highlighted numerous advantages of using smart glasses (head-mounted devices) over smartphones and hand-held devices. A key advantage of smart glasses is that they follow the natural gaze of the user's eyes and head, which eliminates the need for awkward pointing or manoeuvring to capture what is in front of them, as noted by P22:

"Generally people are looking at what they are doing, so it's natural to hold something up to the face, to look at it and feel it, so there by having it there [head-mounted],

having it hands free, having not to worry about holding something."

Smart glasses also offer an additional benefit over hand-held devices in that they can be used without requiring the use of both hands. This is especially advantageous for guide dog users who may already be using one hand to hold the guide dog. P1 noted that:

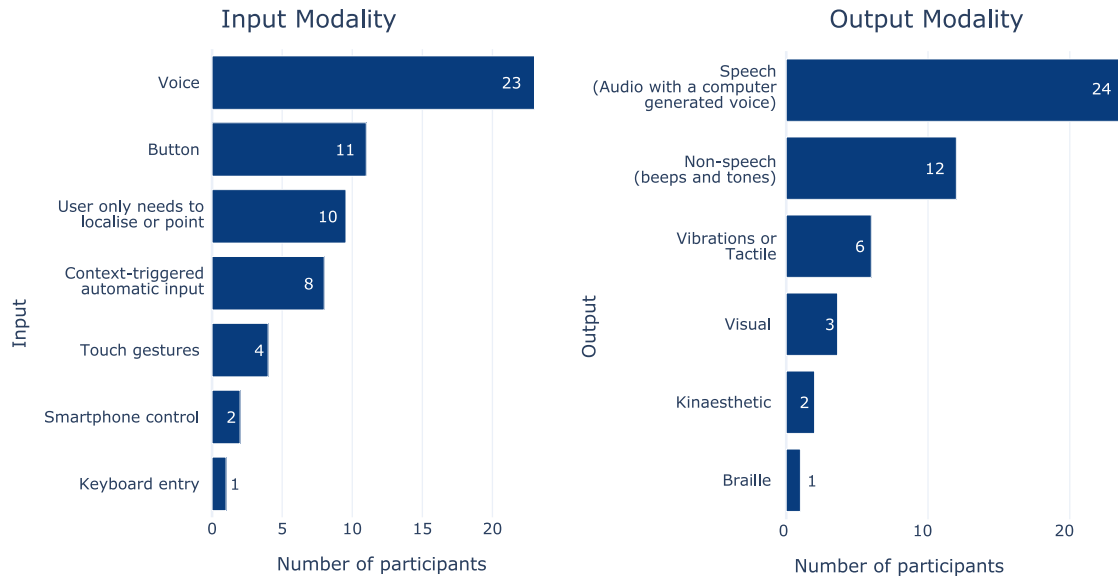


Figure 6: Interaction modes identified in the interview with the number of participants that preferred it for at least one of their top five tasks

I have a guide dog that I'm out and about with, so that means that one of my hands is always busy with the dog.

Finally, our comparison between the blind and low-vision group found no discernible preference in the type of device for their top five tasks.

Ranking of researcher tasks: Figure 7 shows the ranking assigned by participants to the tasks identified in the survey. Image Captioning, Personal Object Recognition, Safe Path Detection and Fill Paper Forms all received the same highest score. Several participants also expressed the need for a smart device that could assist them in filling out paper forms, as they currently have no alternative but to rely on a sighted person. One participant, P4, emphasised the importance of accurate Image Captioning, as it renders them unable to comprehend the content without it.

Appliance Recognition and Scene Simplification were ranked as the lowest in terms of usefulness. One reason why Appliance Recognition was not useful, as pointed out by P4, was that she sets up her living space in a way that allows her to easily identify appliances, making a smart device redundant in this regard. P3 suggested that Scene Simplification may not be useful as he did not want to rely on a device that uses vision, given the knowledge that his vision would deteriorate over time.

In the course of our interview, we made an intriguing observation regarding the tasks extracted from the paper [14], which aimed at developing a system for detecting Animal Waste or Waste to avoid stepping on it. Upon being asked during the interviews, two out of six participants stated that their primary use for the device would be to locate the animal waste left by their guide dog. This is a significant consideration, as visually impaired individuals face challenges in picking up after their guide dogs. They explained

that if the dog defecates outside, they may have no choice but to leave the waste as they are unable to locate it. This suggests that development of a smart device or smartphone application that can detect guide dog waste would greatly benefit guide dog users.

We conducted an analysis to measure if low-vision and blind participants had different task rankings. Figure 8 presents the mean difference in scores assigned by the two groups. Our findings indicated that low-vision participants assigned greater importance to Overlays and Scene-Segmentation, whereas blind participants favoured Barcode & QR Code Reader and Color Detection. The explanation of Overlays and Scene-Segmentation is straightforward for low-vision participants as they can depend on the visual mode of interactions. Blind participants mentioned that a device capable of identifying and reading Barcode and QR codes would be more useful as they are not able to locate them. All four blind participants mentioned that Color Detection is invaluable (3) or very useful (1) especially if it was accurate even under different lighting conditions, compared to the two low-vision participants mentioning it will only be slightly useful. Additionally, our analysis identified that Text Recognition tasks exhibited no significant difference between the two user groups. This, in conjunction with the results from the previous section regarding the participants preferred tasks, highlights the necessity for accurate text recognition devices regard of the visual condition. However, it is important to note that the sample size for this comparison was extremely limited, with only six participants ranking each task and only 2-3 of them being low-vision participants.

Based on the 'best' type of device selected by the participants for each task, we calculated the count of all the wearables as shown in Column C in Figure 10. One thing to note is, participants were allowed to select more than one 'best' wearable for each task. For example, one participant gave hand-held device and smartphone for

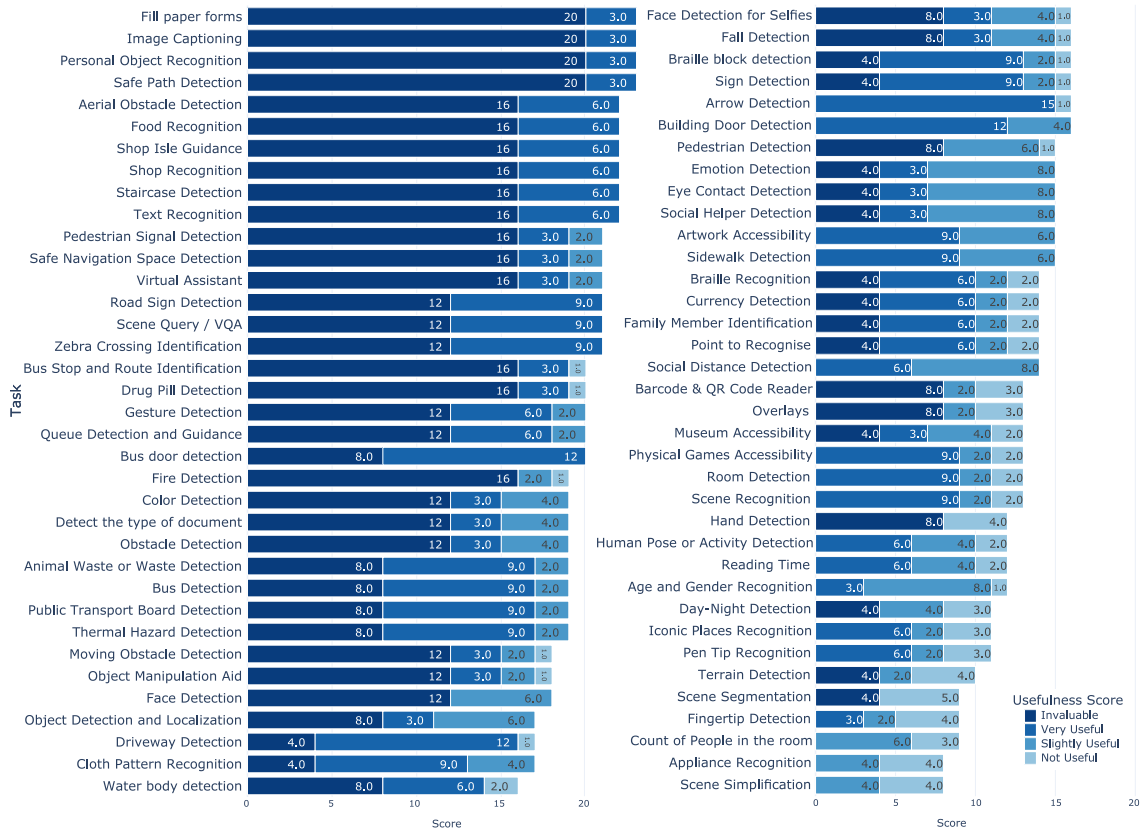


Figure 7: Ranking of tasks identified in the review based on participants' score of usefulness, ranging from "Not Useful," "Slightly Useful," "Very Useful," to "Invaluable". Each category was assigned a score from 1 to 4 and the tasks are then sorted based on the total score. The figure is presented from left to right and top to bottom.

Color detection. Based on this, head-mounted devices, smartphones, and hand-held devices were the top three device types for each of the tasks. We also found no discernible preference in the type of device between the blind and low-vision participants.

Finally, six out of the 24 participants chose to update their top five list when given the opportunity to do so after ranking the tasks.

Overall perception of smart devices: All participants expressed that they considered computer vision based smart devices to be highly valuable for their future needs. P22 summarised this clearly:

"Because of the independence they offer a person who is blind or visually impaired to act as normally as possible. In external environment, it build somebody's confidence and sense of well being. In a home environment, it would mean not relying on others all the time to assist with visual tasks and it would increase safety and save alot of time."

Subsequently, we inquired whether participants had any concerns or design considerations that the research community should take into account while creating AI-enabled smart devices. Of the 24 participants, 16 cited ease of use as their top concern, encompassing factors such as longer battery life, reliability, and portability.

Another 10 participants emphasised the need for smart devices to be subtle and visually appealing. As P5 encapsulated this sentiment:

Nobody wants to walk around in Big Ski Goggles, because you feel conspicuous, and everybody's staring at you for a different reason, not just for being blind, because you're in these great big, ugly glasses.

The majority of these current limitations of smart devices are due to hardware constraints, which were discussed in previous studies [44]. However, it is reasonable to anticipate that these limitations will be addressed as the technology evolves, resulting in devices that are both smaller in size and more powerful.

5 COMPARISON & DISCUSSION

In this section we contrast the results of the two studies and also discuss the implications for smart device research and adoption.

5.1 Which tasks should researchers explore?

We were interested to see the alignment between researcher focus on tasks as revealed by the number of papers addressing the task in Study 1 and the BLV participants' ranking of the importance of the task found in Study 2. We first computed the correlation

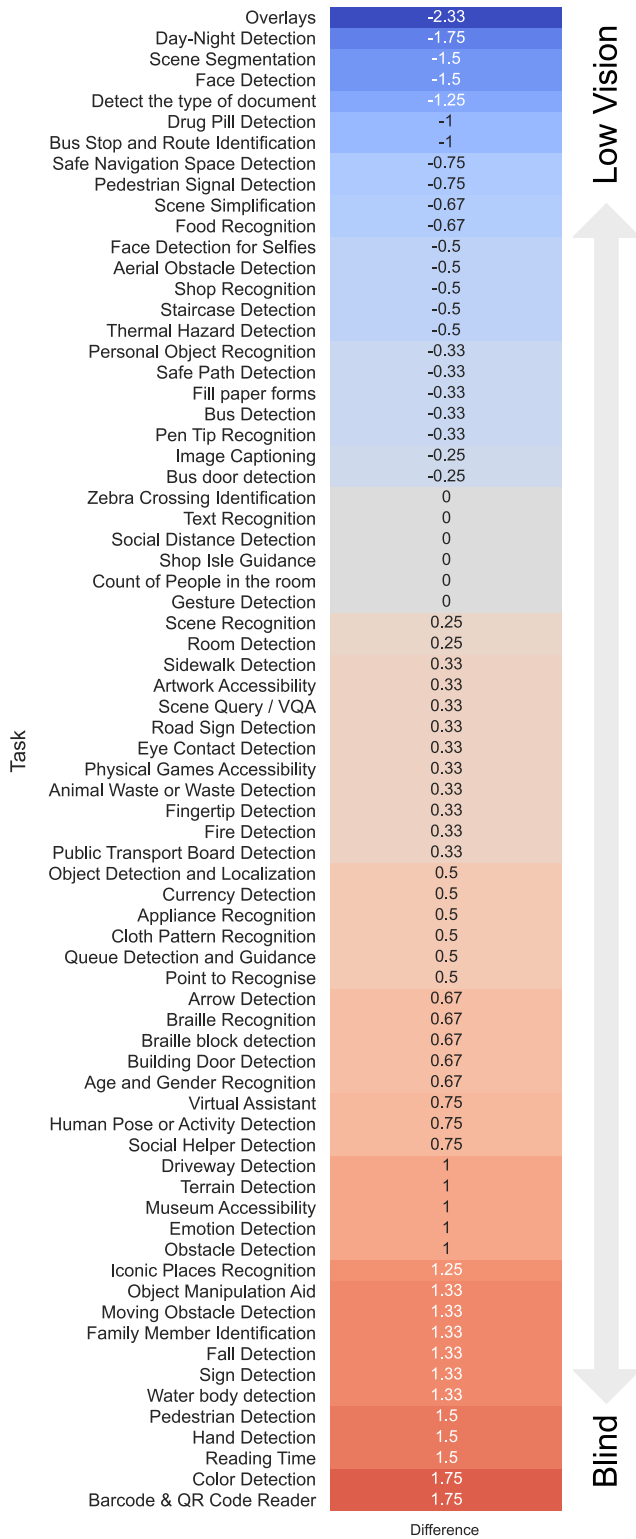


Figure 8: Mean difference of usefulness score of tasks between the blind and low-vision groups.

between the two. This was $r(70) = .14, p=.24$, indicating a weak positive correlation.

We then investigated the differences for each task. The tasks identified in Study 1 were ranked based on the count of studies identified for each task (named "*Review Rank*"). In Study 2, tasks were ranked based on the perceived usefulness of the tasks by the participants (named "*Interview Rank*"). We scaled and plotted these two rankings in Figure 9 to explore differences in the ranking. This reveals:

- Tasks that were **important** to the participants, but were **not the focus** of the researchers (highlighted in Blue with a [+] symbol)
- Tasks that were deemed **unimportant** by the participants, but were the **focus** of the researchers (highlighted in Red with a [-] symbol)

Tasks that were important to the participants, but were not the focus of the researchers: Although filling paper forms had the lowest number of studies, we found that all six participants who ranked it, gave a higher usefulness score (Invaluable (5 participants) and Very Useful (1 participant)). In fact, only one group of researchers have worked on the filling paper forms task, as reported in [26, 40, 41]. Interestingly, these authors found that all of their participants stated they would use such a device in their everyday life. Considering our findings, we suggest researchers should be prioritising tasks that have been highlighted in blue, such as Aerial Obstacle Detection, Shop Recognition, and Safe Navigation Space Detection.

Although we have highlighted few specific tasks based on the existing research, we acknowledge that other tasks may also be significant, as indicated by the bolded text in Figure 5. Participants in our study identified tasks such as Empty Seat Detection, Sorting Items, and Checking Appearance as important. Additionally, researchers can also explore tasks for which BLV users currently rely on sighted assistance, as described in Section 4.3. Therefore, we suggest that researchers consider both existing research and participant feedback when identifying significant tasks for smart devices aimed at aiding the BLV community.

Tasks that were deemed unimportant by the participants, but were the focus of the researchers: Tasks such as Terrain Detection, Currency Detection and Pedestrian Detection were among the most extensively studied in the literature. However, participants reported that they did not consider these tasks to be important. For instance for Terrain Detection, P7 stated that they could identify terrain by the sensation under their feet, while P10 noted that others would warn them about slippery conditions. For Pedestrian Detection, participants who used guide dogs indicated that the dogs are good at avoiding collisions, while others reported that they could sense when someone was approaching through hearing or tactile cues. Participant P24 offered additional insight into this:

"I think as a cane user that would be useful, [...] because I'm a dog user, it would actually be not so useful because the dog detect those people and walks around them, or people see you coming, and they might move out of the way."

It is also important that these devices need to gain the trust of the BLV community as described by [81]. So smart devices should be

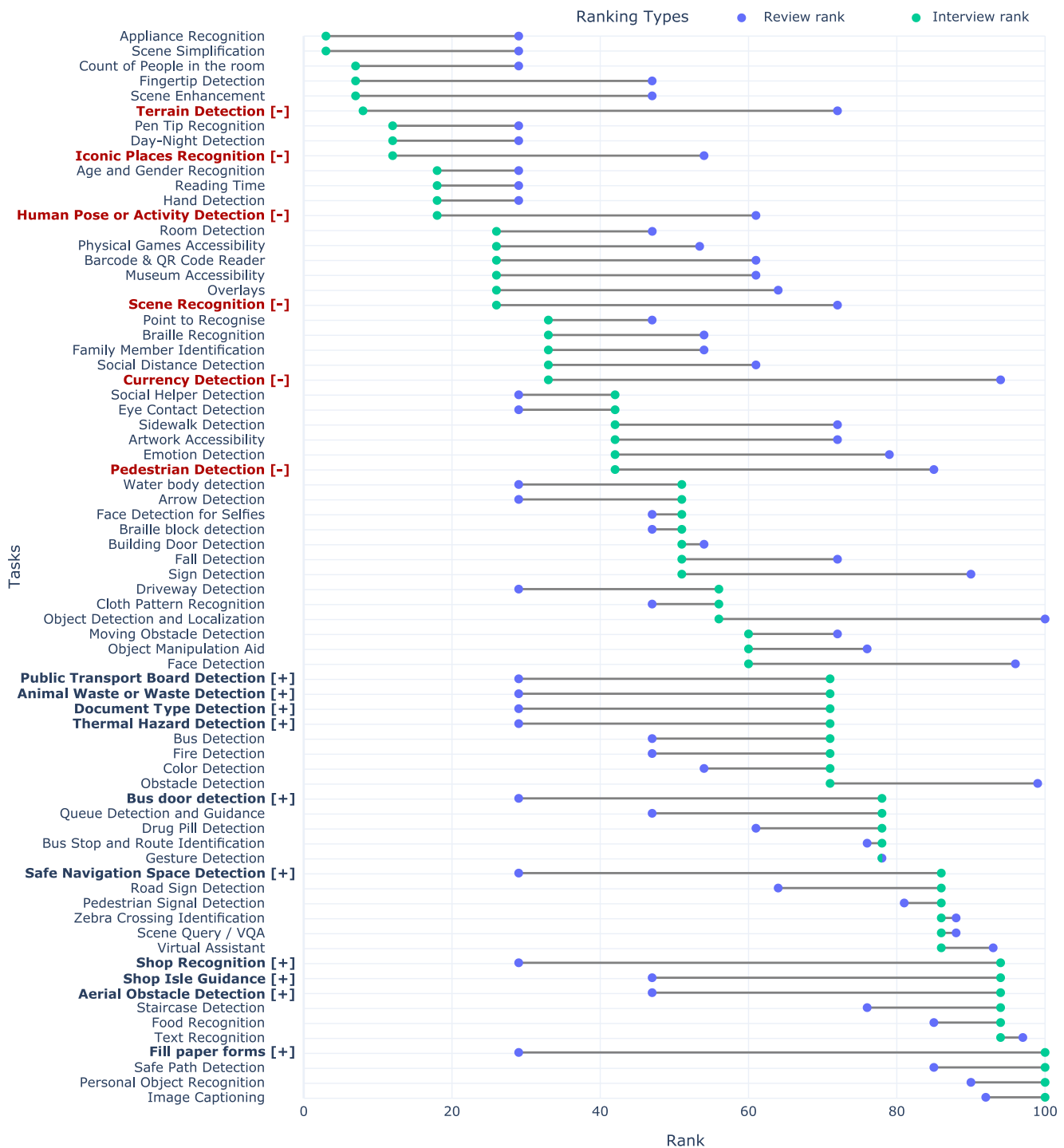


Figure 9: Ranking of various tasks based on the number of papers in the review [Review rank] and interview preference [Interview rank]. Tasks that were important to the participants, but were not the focus of the researchers are highlighted in blue with a [+] symbol next to them and tasks that were deemed unimportant by the participants, but were the focus of the researchers are highlighted in red with a [-] symbol next to them.

designed to work in tandem with the currently established orientation and mobility standards, particularly as these technologies continue to evolve and mature.

Our analysis also revealed that Currency Detection exhibited the greatest disparity between the literature and the perceptions of our participants. However, it is important to note that the majority of our participants were from Australia, a country where (in particular since the pandemic) cashless transactions are the norm, rendering the ability to identify individual notes and coins less critical. Furthermore Australian currency has been designed to be tactually distinct. To gain a more comprehensive understanding of the importance of Currency Detection among individuals who rely heavily on cash transactions, future studies should consider recruiting participants from regions where cash remains a primary form of payment and the currency may be difficult to distinguish tactually.

5.2 Which interaction modalities are preferred?

Question Answering interaction model as a conversational agent: All 24 of our participants expressed the desire for at least one of their top tasks to use a Question Answering (QA) interaction model. This is in line with the increasing popularity of voice input and speech output modalities among BLV participants [20, 83], and emphasises the importance of smart devices acting as conversational agents. It also accords with social agency theory [72, 74, 85], which suggests that users become more engaged when they are presented with a human-human social interaction.

A conversational interface is particularly suited to more complex tasks such as the second most requested task (Figure 5), "Scene Query" (9 participants), in which users wish to ask the device questions they might have about their surroundings. As P8 expressed, when they want to understand the menu at a restaurant, they prefer to engage in a natural conversation with the device to understand the items on it, as opposed to simply having it read text off the menu. The ideal device would function as a conversational agent, capable of switching contexts and serving as a personal companion that users can trust.

Since 2010 [25], the field of question answering for blind people has seen significant exploration. This includes the collection of large datasets of visual questions and answers such as VizWiz [47]. However, our review only identified 13 studies that employed the Question Answering interaction model in smart wearable devices suggesting that this area remains under-explored in this context.

We also note that the emergence of GPT-4 [77] has led to the introduction of a beta version of Envision glasses, known as "Ask Envision" [7], aiming to provide seamless user-device interactions through conversational interfaces. This potentially enables question answering functionalities on wearable devices.

Another crucial aspect of the interaction is determining the appropriate amount and timing of information delivery to the user. A recurring concern among participants was not to be overwhelmed by information and they expressed a desire for devices to only share information that they find interesting/relevant. For instance, participant P14 stated, "In an ideal world, you want it [smart device] to read only what I'm interested in."

We also identified that the devices should present information to the user in a manner that is easily understandable, such as using step-based distances instead of meters and feet, and directional indicators such as "to the right" or "at 2 o'clock" instead of cardinal directions like North, East, South, West. As participant P1 stated:

"If a device was giving me direction, that [in form of steps] would be something I would understand easily. [...] If somebody said five meters, I would have to kind of process that, but if somebody said 10 steps that would be more efficient information for me."

The majority of studies relied on automated text-to-audio tools to read out the detected item [19, 33, 100]. Therefore, further exploration is needed into this issue, as only a limited number of studies have focused on how information is conveyed to the user [70, 71].

Finally, two participants expressed their preference for the device not to read sensitive information aloud or cause disturbance to those around them. This issue has also been identified in previous studies where the majority of participants expressed concerns about privacy in speech-based smart devices [53, 99]. To address this issue, two potential solutions can be considered: desensitising the device when reading sensitive information aloud or using bone-conducting headphones to ensure privacy.

How important is the visual modality for low-vision individuals?: Several studies have investigated the potential of image enhancement as an assistive interface for people with low-vision to perceive and comprehend their surroundings [42, 101, 103, 104]. However, our study found that our low-vision participants did not prefer visual feedback, and only 3 (out of the 10 low-vision) participants mentioned they would prefer visual output. We found this surprising and it may be because most of our low-vision participants had only light perception and progressive visual impairments, as discussed in the limitations section.

In a 2020 study by Zhao et al. [102], they compared the cognitive load between visual and audio feedback during navigation and found that 4 out of the 16 participants reported that the visual feedback was more distracting due to the extra effort required to perceive and understand the visual content, compared to following audio instructions. While this finding accords with our findings, we believe that other factors, such as environment (crowded, calm), amount of information, and the specific visual condition, also have an impact on preference. Therefore, further research is needed to clearly understand the factors that lead to the preference for visual or audio modalities.

5.3 Which devices are preferred?

Figure 10 reveals large differences between the preferences of BLV participants in Study 2 and the devices used by researchers in Study 1. We see that researchers have focused on smartphone applications, body-mounted devices and then head-mounted displays. However body-mounted devices were rarely mentioned by participants and participants had a strong preference for head-mounted devices over other devices, even if we aggregate hand-held devices and smartphones into a single category.

The authors of [99] compared the benefits of wearables compared to smartphones and also found that head-mounted devices have the most potential to provide multiple interaction modalities for

Wearable	Ankle-Worn Device	2		
	Audio headset [+]		4	1
	Backpack	4		
	Belt	7	2	
	Bracelet [+]			3
	Body-mounted Device	187		10
	Cane	39	1	32
	Cap	13		
	Chest-mounted Device	14	8	22
	Contact lens	1	1	5
	Finger-worn Device	3		1
	Glove	10		9
	Guidedog-mounted Device [+]			4
	Hand-held Device	16	19	68
	Hand-worn Device	6		3
	Head-mounted Device	157	70	258
	Hoodie	1		
	Keychain [+]			1
	Mirror [+]			1
	Necklace	1	4	7
	Pen [+]			1
	Pendent [+]		2	1
	Ring [+]		1	3
	Robot	4	10	18
	Robotic Arm	1		4
	Shoes	10	1	
	Smart Camera	1		
	Smart Watch	2	5	60
	Smartphone	215	28	75
	Suitcase	1		1
	Walker	1		
	Wrist-worn Device	2		7
		A	B	C
		Counts		

Figure 10: Comparison of Wearable Device Type Counts. (A) Device counts across studies in the review. (B) Count of devices preferred by participants for their top 5 tasks. (C) Counts of devices preferred by participants for different tasks in the review. Devices not found in our review are indicated in bold font with a [+] symbol next to them.

better accessibility. This trend aligns with the growing use of head-mounted devices as an assistive tool for the visually impaired, as identified by Li et al. (2022) [65].

Additionally, we identified wearables that were not included in the review but were mentioned by the participants (highlighted in bold in Figure 10). One notable wearable type is guide dog mounted wearables, as individuals who rely on guide dogs could potentially use the harness to mount a smart device on it. Although no studies were found in the review that utilised such a system, it may be a promising avenue for future research.

We have also observed that the selection of a wearable device depends on the task and the environment in which it is being performed. P16 mentioned that using a hand-held device or smartphone in public can attract unwanted attention and increase the risk of theft. In contrast, a smart glass can blend in as a regular pair of glasses and reduce the risk of snatch and runs (as participants mentioned in this study [99]). However, when at home, a hand-held device or smartphone app would be more appropriate as they are more aware of their surroundings and have both hands free. Other participants also indicated a preference for hand-held devices for color detection, which is mainly used at home, while smart glasses were preferred for text recognition, which is used in both public and home settings. Therefore, it appears that the choice of wearable device could be determined by the specific task and the context in which it is used.

Universal Device or Platform: The results of our study showed that ease of use is a top concern for blind and low-vision users when it comes to smart devices. This is consistent with findings from previous studies [69, 86, 99]. One noteworthy observation was that three participants expressed the opinion that if a particular task can already be performed using an existing device, such as a smartphone, there is no need to develop a separate device solely for that task. This sentiment was voiced by participant P4:

"We already have existing smart devices, so let's reduce the design for all these new devices that are ridiculously priced, people can't afford, and we end up walking around which so much stuff hanging off us."

This observation highlights a significant challenge that may face real-world deployment of smart devices. Our review revealed that much of the research in this field is conducted in isolated silos, resulting in the creation of separate wearable devices and applications for different tasks. Implicitly, this means that users are expected to switch between these devices as needed, which will be inconvenient and they may also need to purchase multiple devices.

We believe there is a pressing need to develop only a few multipurpose devices and applications that seamlessly work together on these devices. In our ideal world, a user with visual impairment will be able to utilise a single smart assistive device (probably head-mounted) to gain access to information regarding buses, safely board the bus, locate unoccupied seats, carry out shopping at their intended destination, and participate in social interactions with peers without experiencing any form of stigma.

Finally, an essential aspect highlighted by five participants was the need for adequate training and opportunities to test these devices before making a purchase. Given the diverse range of visual conditions that these devices must accommodate, it is advisable to provide potential buyers with the opportunity to test them before making a purchase. As P2 said:

"Instead of saying, oh! this [smart-device] is available, go and buy it, they should ask someone that either can come to your home or you can go to the organisation, and have a look at it and [see] if you feel comfortable with it. I'm sure they have things that they can lend out, so that you know to see if it's going to be suited to your needs, because some things aren't suited to your needs, and unless you use them you don't know, because you're not experiencing it. It also makes you feel more confident when you have bought it, because you know it's going to work for you."

Tapu et al. [92] also noted this, highlighting the absence of structured training protocols for individuals with visual impairment when using smart assistive devices. The authors noted that BLV users often have to rely on self-directed learning without access to a trainer or clear set of instructions.

5.4 Why is Blind and Low-Vision community involvement important in smart assistive device research?

As identified by this study, one of the major challenges encountered in current research into smart devices is the mismatch between researcher choice of task and device and the preferences of BLV people. We believe that at least in part this problem arises because of a lack of involvement of BLV people in the research, and, in particular, involvement in the design stage.

To investigate if this is the case, we computed the correlation coefficient between the usefulness scores given by the participants and count of papers categorised from the review as follows:

Studies	Correlation Coefficient
No BLV participants	$r(70)=.13, p=.28$.
With BLV participants only involved in the evaluation stage	$r(70)=.16, p=.18$.
With BLV participants involved in the design stage	$r(70)=.27, p=.02$.

This reveals that studies that included BLV participants during the design stage demonstrated a significantly higher positive correlation compared to the other two groups. While correlation does not necessarily mean causation, this does suggest that involving BLV participants in the design stage may lead to the development of smart devices that perform useful tasks for the community.

Nonetheless, it should be noted that overall, there is no strong correlation indicating that BLV participant involvement guarantees the creation of a device that is generally perceived to be useful by the BLV community. We believe identifying the most useful applications of new assistive technologies is best served by studies like this and that of [44] that encourage BLV participants to "brainstorm" possible applications of these technologies.

A number of other studies have demonstrated the advantages of involving users in the development process, including but not limited to, generating superior ideas, shortening the development cycle, and increasing user satisfaction [23, 59, 90]. Duckett & Pratt [39] also found that the BLV community believes that researchers

could make a greater contribution to the field by consulting them more thoroughly in research design and practice, and by giving them more control over research activities.

Our findings suggest that many smart-device studies prioritise development of the technology over the requirements of the user, which could potentially lead to a low adoption rate of these devices and a disconnect between the research community and the intended user base [87].

We believe it is imperative that researchers recognise the importance of involving BLV participants from the beginning of the research process through exploratory studies and adopting a co-design approach. However, we note that involving BLV participants can be challenging due to various factors, such as the difficulty in finding willing participants, ensuring appropriate reimbursement, and addressing ethical considerations [28].

Another consideration was raised by [65] where they showed that many of the studies on head-mounted-devices were done in a controlled lab environment thereby not reflecting the actual real world use cases. So more studies involving BLV participants also needs to be done in read world scenarios.

6 LIMITATIONS

Our study had a limited sample size (24 participants), and only six participants were asked to rank the tasks identified in the review, and only two to three of these had low-vision and only three or four identified as blind. This may not be sufficient to draw conclusive results, in particular differences between preferences of participants who are blind and those with low vision, and further studies with larger sample sizes are warranted. Additionally, the low-vision participants in our study had, on average, visual acuity closer to total vision loss with light perception and had progressive vision loss. This may have impacted the results, in particular the low preference for visual output.

Another limitation is that our study sample was predominantly drawn from a single developed country, with only one participant from a developing country. Moreover, the participants were recruited through an electronic mailing list, which may have resulted in a self-selection bias, with individuals who are more technology-savvy being over represented in our sample. These factors are reflected in the technology adoption rates of our participants, as presented in Table 4 (in the Appendix).

We acknowledge that our study was limited to research conducted between 2020 and mid 2022. Consequently, if there are significant bodies of works prior to or beyond this period they will not have been included in our analysis.

It is also crucial to acknowledge the rapidly evolving field of smart devices. With the introduction of GPT models, such as GPT4 [77], devices will soon be capable of capturing visual scenes and answering users' prompted questions. This is exemplified in the new features by Envision [7] and Be My Eyes [9]. Our study did not explicitly consider how BLV users would like to interact with GPT4-enabled devices and further studies are required.

7 CONCLUSION

In this paper, we conducted two studies to evaluate whether current smart assistive technology research utilising AI-based computer vision to understand the immediate environment addresses the needs of blind and low-vision (BLV) users. Study 1 involved a scoping review of 646 papers, while Study 2 comprised semi-structured interviews with 24 BLV participants to ascertain their preferences.

Our findings reveal only a weak positive correlation between the perceived importance of tasks by BLV users and researchers' focus on those tasks. Although BLV participant involvement during the research design stage improved the correlation slightly, it remained low. Moreover, conversational agent type devices are preferred by BLV participants, but limited research has been conducted on this topic. We found that head-mounted devices are preferred by BLV users, although the choice of wearable device depends on the task and environment, and that there was only limited preference for visual output by our low-vision participants. Finally, we suggest that researchers focus on more universal device or platform rather than bespoke devices so as to enable seamless usage without the need for device switching.

We believe this research will provide valuable guidance for researchers working in the field of smart assistive technologies and enable them to focus on tasks and interactions that are prioritised by the BLV community.

ACKNOWLEDGMENTS

We are grateful to the Monash Data Futures Institute for providing the funding that made this project possible. We would like to extend our appreciation to our participants Adua Merola, Cass Embling, Cynthia Kate Gregory, Debra L Simons, Duc Anh Minh Nguyen, Elise Lonsdale, Erin Goedhart, Grace King, Iris Wilson, Janene Sadhu, Jennifer Parry, John Hardie, Josephine Mckinnon, Leonore Scott, Maree Fenech, Marée Steinway, Meredith Prain, Michael Hamon, Neslihan Sari, Peter Mason, Rocco Cutri, Susie Rich, Yvonne Huntley and other participants for their invaluable contribution to our study. Their active participation significantly enriched the study, and we are truly appreciative of their involvement.

REFERENCES

- [1] 2011. Assistive products for persons with disability – Classification and terminology. <https://www.iso.org/standard/50982.html>
- [2] 2015. AS/NZS ISO 9999:2011 Assistive products for persons with disability—classification and terminology. <https://meteor.aihw.gov.au/content/508624>
- [3] 2023. Be my eyes - see the world together. <https://www.bemyeyes.com/>
- [4] 2023. Empowering your vision. <https://www.nueyes.com/>
- [5] 2023. Envision glasses. <https://www.letsenvision.com/glasses>
- [6] 2023. Google lookout. <https://play.google.com/store/apps/details?id=com.google.android.apps.accessibility.reveal>
- [7] 2023. Introducing 'Ask Envision': A Gamechanger for the blind and low vision community. <https://www.letsenvision.com/blog/introducing-ask-envision-a-gamechanger-for-the-blind-and-low-vision-community>
- [8] 2023. Introducing ChatGPT. <https://openai.com/blog/chatgpt>
- [9] 2023. Introducing our virtual volunteer tool for people who are blind or have low vision, powered by OpenAI's GPT-4. <https://www.bemyeyes.com/blog/introducing-be-my-eyes-virtual-volunteer>
- [10] 2023. Seeing AI app from Microsoft. <https://www.microsoft.com/en-us/ai/seeing-ai>
- [11] 2023. Strap tech. <https://strap.tech/home>
- [12] 2023. Vision buddy - television watching for visually impaired. <https://visionbuddy.com/>
- [13] 2023. Visual interpreting – get live, on-demand access to visual information. <https://aira.io/>
- [14] R Abirami, S Haarini, and P Hari. 2020. Customized Smart Glasses for Needy Blind People. (2020).
- [15] Younes Akbari, Hanadi Hassen, Nandhini Subramanian, Jayakanth Kunhoth, Somaya Al-Maadeed, and Wael Alhajjaseen. 2020. A vision-based zebra crossing detection method for people with visual impairments. In *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*. IEEE, 118–123.
- [16] Mojisola Rachael Akinsiku, Brendan Ubochi, and Waliu Apena. 2020. Smart Walking Stick for the Blind with Text Image Recognition. *Journal of Electrical Engineering, Electronics, Control and Computer Science* 6, 4 (2020), 17–22.
- [17] Erick Javier Argüello Prada and Lina Maria Santacruz Forero. 2022. A belt-like assistive device for visually impaired people: Toward a more collaborative approach. *Cogent Engineering* 9, 1 (2022), 2048440.
- [18] Adwitiya Arora, Atul Grover, Raksha Chugh, and S Sofana Reka. 2019. Real time multi object detection for blind using single shot multibox detector. *Wireless Personal Communications* 107 (2019), 651–661.
- [19] Fahad Ashiq, Muhammad Asif, Maaz Bin Ahmad, Sadia Zafar, Khalid Masood, Toqeer Mahmood, Muhammad Tariq Mahmood, and Ik Hyun Lee. 2022. CNN-based object recognition and tracking system to assist visually impaired people. *IEEE Access* 10 (2022), 14819–14834.
- [20] Shiri Azenkot and Nicole B. Lee. 2013. Exploring the Use of Speech Input by Blind People on Mobile Devices. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility* (Bellevue, Washington) (ASSETS '13). Association for Computing Machinery, New York, NY, USA, Article 11, 8 pages. <https://doi.org/10.1145/2513383.2513440>
- [21] Fereshteh S Bashiri, Eric LaRose, Jonathan C Badger, Roshan M D'Souza, Zeyun Yu, and Peggy Peissig. 2018. Object detection to assist visually impaired people: A deep neural network adventure. In *Advances in Visual Computing: 13th International Symposium, ISVC 2018, Las Vegas, NV, USA, November 19–21, 2018, Proceedings 13*. Springer, 500–510.
- [22] George M Beal, Joe M Bohlen, et al. 1957. *The diffusion process*. Agricultural Experiment Station, Iowa State College.
- [23] Cynthia L. Bennett and Daniela K. Rosner. 2019. The Promise of Empathy: Design, Disability, and Knowing the "Other". In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, 1–13. <https://doi.org/10.1145/3290605.3300528>
- [24] Carlos Bermejo, Tristan Braud, Ji Yang, Shayan Mirjafari, Bowen Shi, Yu Xiao, and Pan Hui. 2020. VIMES: A Wearable Memory Assistance System for Automatic Information Retrieval. In *Proceedings of the 28th ACM International Conference on Multimedia* (Seattle, WA, USA) (MM '20). Association for Computing Machinery, New York, NY, USA, 3191–3200. <https://doi.org/10.1145/3394171.3413663>
- [25] Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samuel White, and Tom Yeh. 2010. VizWiz: Nearly Real-Time Answers to Visual Questions. In *Proceedings of the 23rd 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>
- [26] Syed Masum Billah, Vikas Ashok, and IV Ramakrishnan. 2018. Write-it-yourself with the aid of smartwatches: A wizard-of-oz experiment with blind people. In *23rd International Conference on Intelligent User Interfaces*. 427–431.
- [27] Alain Boldini, John-Ross Rizzo, and Maurizio Porfiri. 2020. A piezoelectric-based advanced wearable: obstacle avoidance for the visually impaired built into a backpack. In *Nano-, Bio-, Info-Tech Sensors, and 3D Systems IV*, Vol. 11378. SPIE, 1137806.
- [28] Emeline Brulé, Brianna J. Tomlinson, Oussama Metatla, Christophe Jouffrais, and Marcos Serrano. 2020. Review of Quantitative Empirical Evaluations of Technology for People with Visual Impairments. 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/10.1145/3313831.3376749>
- [29] Diego Renan Bruno, Marcelo Henrique de Assis, and Fernando Santos Osório. 2019. Development of a mobile robot: Robotic guide dog for aid of visual disabilities in urban environments. In *2019 Latin American Robotics Symposium (LARS), 2019 Brazilian Symposium on Robotics (SBR) and 2019 Workshop on Robotics in Education (WRE)*. IEEE, 104–108.
- [30] Stacy E Butler, Adele Crudden, William K Sansing, and BJ LeJeune. 2002. Employment barriers: Access to assistive technology and research needs. *Journal of Visual Impairment & Blindness* 96, 9 (2002), 664–667.
- [31] Arlene R. Caballero, Kristopher Elijah I. Catli, and Aldwin Garry F. Babierra. 2020. Object Recognition and Hearing Assistive Technology Mobile Application Using Convolutional Neural Network. In *Proceedings of the 2020 International Conference on Wireless Communication and Sensor Networks* (Warsaw, Poland) (icWCSN '20). Association for Computing Machinery, New York, NY, USA, 41–48. <https://doi.org/10.1145/3411201.3411208>
- [32] Qingqing Cao, Prerna Khanna, Nicholas D. Lane, and Aruna Balasubramanian. 2022. MobiVQA: Efficient On-Device Visual Question Answering. *Proc. ACM*

- Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 44 (jul 2022), 23 pages. <https://doi.org/10.1145/3534619>
- [33] Wan-Jung Chang, Liang-Bi Chen, Chia-Hao Hsu, Jheng-Hao Chen, Tzu-Chin Yang, and Cheng-Pei Lin. 2020. MedGlasses: A wearable smart-glasses-based drug pill recognition system using deep learning for visually impaired chronic patients. *IEEE Access* 8 (2020), 17013–17024.
- [34] Ching-Han Chen and Ming-Fang Shiu. 2020. Smart guiding glasses with descriptive video service and spoken dialogue system for visually impaired. In *2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan)*. IEEE, 1–2.
- [35] Roe Cohen. 2023. Experience the power of assistive technology with Ocrecam's AI devices. <https://www.orecam.com/en-us/home>
- [36] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.
- [37] Pei Du and Nirupama Bulusu. 2021. An Automated AR-Based Annotation Tool for Indoor Navigation for Visually Impaired People. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility (Virtual Event, USA) (ASSETS '21)*. Association for Computing Machinery, New York, NY, USA, Article 50, 4 pages. <https://doi.org/10.1145/3441852.3476561>
- [38] Pei Du and Nirupama Bulusu. 2021. An automated AR-based annotation tool for indoor navigation for visually impaired people. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–4.
- [39] Paul S. Duckett and Rebekah Pratt. 2001. The Researched Opinions on Research: Visually impaired people and visual impairment research. *Disability & Society* 16, 6 (2001), 815–835. <https://doi.org/10.1080/09687590120083976> arXiv:<https://doi.org/10.1080/09687590120083976>
- [40] 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>
- [41] Shirin Feiz, Anatoliy Borodin, Xiaojun Bi, and IV Ramakrishnan. 2021. Towards Enabling Blind People to Fill Out Paper Forms with a Wearable Smartphone Assistant. In *Proceedings. Graphics Interface (Conference)*, Vol. 2021. NIH Public Access, 156.
- [42] Dylan R Fox, Ahmad Ahmadzade, Clara Tenia Wang, Shiri Azenkot, Marlena A Chu, Roberto Manduchi, and Emily A Cooper. 2023. Using augmented reality to cue obstacles for people with low vision. *Optics Express* 31, 4 (2023), 6827–6848.
- [43] Omar Gamal, Shubham Thakkar, and Hubert Roth. 2020. Towards Intelligent Assistive System for Visually Impaired People: Outdoor Navigation System. In *2020 24th International Conference on System Theory, Control and Computing (ICSTCC)*. 390–397. <https://doi.org/10.1109/ICSTCC50638.2020.9259682>
- [44] Eugenie Golubova, Sandra D Starke, Michael D Crossland, and James S Wolffsohn. 2021. Design considerations for the ideal low vision aid: insights from de-brief interviews following a real-world recording study. *Ophthalmic and Physiological Optics* 41, 2 (March 2021), 266–280. <https://doi.org/10.1111/opo.12778>
- [45] Ravi Gorapudi, Puvvula Priya Darsini, Uddanti Naga Kavya, and Oggur Jaswanthi. 2020. Product label, obstacle and sign boards detection for visually impaired people. In *International Conference of Advance Research & Innovation (ICARI)*. 46.
- [46] Renan Guarese, Franklin Bastidas, João Becker, Mariane Giambastiani, Yhonatan Iquiapaza, Lennon Macedo, Luciana Nedel, Anderson Maciel, Fabio Zambetta, and Ron van Schyndel. 2021. Cooking in the Dark: Exploring Spatial Audio as MR Assistive Technology for the Visually Impaired. In *IFIP Conference on Human-Computer Interaction*. Springer, 318–322.
- [47] 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. arXiv:1802.08218 [cs.CV]
- [48] Paul J Hershberger. 1992. Information loss: The primary psychological trauma of the loss of vision. *Perceptual and motor skills* 74, 2 (1992), 509–510.
- [49] Bin Hong, Zhangxi Lin, Xin Chen, Jing Hou, Shunya Lv, and Zhendong Gao. 2022. Development and application of key technologies for Guide Dog Robot: A systematic literature review. *Robotics and Autonomous Systems* (2022), 104104.
- [50] Jonggi Hong, Christine Vaing, Hernisa Kacorri, and Leah Findlater. 2020. Reviewing Speech Input with Audio: Differences between Blind and Sighted Users. *ACM Trans. Access. Comput.* 13, 1, Article 2 (apr 2020), 28 pages. <https://doi.org/10.1145/3382039>
- [51] MT Islam, MA Rashid, M Ahmad, A Kuwana, and H Kobayashi. 2020. Design and Implementation of Smart Guided Glass for Visually Impaired People. *International Journal of Electrical and Computer Engineering (IJECE)* 9, 4 (2020).
- [52] Rakesh Chandra Joshi, Saumya Yadav, and Malay Kishore Dutta. 2020. YOLO-v3 based currency detection and recognition system for visually impaired persons. In *2020 International Conference on Contemporary Computing and Applications (IC3A)*. IEEE, 280–285.
- [53] Shaun K. Kane, Chandrika Jayant, Jacob O. Wobbrock, and Richard E. Ladner. 2009. Freedom to Roam: A Study of Mobile Device Adoption and Accessibility for People with Visual and Motor Disabilities. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility (Pittsburgh, Pennsylvania, USA) (ASSETS '09)*. Association for Computing Machinery, New York, NY, USA, 115–122. <https://doi.org/10.1145/1639642.1639663>
- [54] M Kathiravan et al. 2021. A Smart Hoodie-Assistive Technology for Blind People Using IoT. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12, 10 (2021), 7565–7569.
- [55] Seita Kayukawa, Tatsuya Ishihara, Hironobu Takagi, Shigeo Morishima, and Chieko Asakawa. 2020. Guiding Blind Pedestrians in Public Spaces by Understanding Walking Behavior of Nearby Pedestrians. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 3, Article 85 (sep 2020), 22 pages. <https://doi.org/10.1145/3411825>
- [56] Akif Khan and Shah Khushro. 2021. An insight into smartphone-based assistive solutions for visually impaired and blind people: issues, challenges and opportunities. *Universal Access in the Information Society* 20 (2021), 265–298.
- [57] Jaejoon Kim. 2020. Application on character recognition system on road sign for visually impaired: Case study approach and future. *International Journal of Electrical and Computer Engineering (IJECE)* 10, 1 (2020), 778–785.
- [58] Matshehla Konaite, Pius A Owolawi, Temitope Mapayi, Vusi Malele, Kehinde Odeyemi, Gbolahan Aiyetoro, and Joseph S Ojo. 2021. Smart Hat for the blind with Real-Time Object Detection using Raspberry Pi and TensorFlow Lite. In *Proceedings of the International Conference on Artificial Intelligence and its Applications*. 1–6.
- [59] Sari Kujala. 2003. User involvement: a review of the benefits and challenges. *Behaviour & information technology* 22, 1 (2003), 1–16.
- [60] Y R Sanjay Kumar, T Nivethetha, P Priyadharshini, and U. Jayachandiran. 2022. Smart Glasses for visually impaired people with facial recognition. In *2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)*. 1–4. <https://doi.org/10.1109/IC3IoT53935.2022.9768012>
- [61] Bineeth Kuriakose, Raju Shrestha, and Frode Eika Sandnes. 2020. Multimodal Navigation Systems for Users with Visual Impairments—A Review and Analysis. *Multimodal Technologies and Interaction* 4, 4 (2020), 73.
- [62] Bineeth Kuriakose, Raju Shrestha, and Frode Eika Sandnes. 2021. SceneRecog: a deep learning scene recognition model for assisting blind and visually impaired navigate using smartphones. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2464–2470.
- [63] Florian Lang, Albrecht Schmidt, and Tonja Machulla. 2020. Augmented reality for people with low vision: symbolic and alphanumeric representation of information. In *Computers Helping People with Special Needs: 17th International Conference, ICCHP 2020, Lecco, Italy, September 9–11, 2020, Proceedings, Part I* 17. Springer, 146–156.
- [64] Chan-Su Lee, Jae-Ik Lee, and Han Eol Seo. 2021. Deep Learning Based Mobile Assistive Device for Visually Impaired People. In *2021 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*. IEEE, 1–3.
- [65] Yifan Li, Kangsoo Kim, Austin Erickson, Nahal Norouzi, Jonathan Jules, Gerd Bruder, and Gregory F Welch. 2022. A Scoping Review of Assistance and Therapy with Head-Mounted Displays for People Who Are Visually Impaired. *ACM Transactions on Accessible Computing (TACCESS)* (2022).
- [66] Jyun-You Lin, Chi-Lin Chiang, Meng-Jin Wu, Chih-Chiung Yao, and Ming-Chiao Chen. 2020. Smart Glasses Application System for Visually Impaired People Based on Deep Learning. In *2020 Indo-Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN)*. IEEE, 202–206.
- [67] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. 2014. Microsoft COCO: Common Objects in Context. <http://arxiv.org/abs/1405.0312> Comment: 1) updated annotation pipeline description and figures; 2) added new section describing datasets splits; 3) updated author list.
- [68] Jack M Loomis, James R Marston, Reginald G Golledge, and Roberta L Klatzky. 2005. Personal guidance system for people with visual impairment: A comparison of spatial displays for route guidance. *Journal of visual impairment & blindness* 99, 4 (2005), 219–232.
- [69] Marie-Céline Lorenzini and Walter Wittich. 2020. Factors related to the use of magnifying low vision aids: a scoping review. *Disability and rehabilitation* 42, 24 (2020), 3525–3537.
- [70] Jun-Li Lu, Hirokyu Osone, Akihisa Shitara, Ryo Iijima, Bektur Ryskeldiev, Sayan Sarcar, and Yoichi Ochiai. 2021. Personalized Navigation that Links Speaker's Ambiguous Descriptions to Indoor Objects for Low Vision People. In *Universal Access in Human-Computer Interaction. Access to Media, Learning and Assistive Environments: 15th International Conference, UAHCI 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part II*. Springer, 412–423.
- [71] Madalin Matei, Lenuța Alboia, and Adrian Iftene. 2022. Safety Navigation using a Conversational User Interface For Visually Impaired People. *Procedia Computer Science* 207 (2022), 1164–1173.
- [72] Richard E Mayer, Kristina Sobko, and Patricia D Mautone. 2003. Social cues in multimedia learning: Role of speaker's voice. *Journal of educational Psychology* 95, 2 (2003), 419.
- [73] Sparsh Mittal and Ankush Mittal. 2011. Versatile question answering systems: seeing in synthesis. *International journal of intelligent information and database*

- systems 5, 2 (2011), 119–142.
- [74] Roxana Moreno, Richard E Mayer, Hiller A Spires, and James C Lester. 2001. The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and instruction* 19, 2 (2001), 177–213.
 - [75] Zachary Munn, Micah DJ Peters, Cindy Stern, Catalin Tufanaru, Alexa McArthur, and Edoardo Aromataris. 2018. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC medical research methodology* 18 (2018), 1–7.
 - [76] T Lavanya Narayani, M Sivapalanirajan, B Keerthika, M Ananthi, and M Arunarani. 2021. Design of Smart Cane with integrated camera module for visually impaired people. In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*. IEEE, 999–1004.
 - [77] OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]
 - [78] Oihana Otaegui, J Seybold, J Spiller, A Marconi, R Olmedo, and M Dubielzig. 2013. ARGUS: Assisting personal Guidance System for people with visual impairment. In *Proceedings of the 26th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2013)*. 2276–2283.
 - [79] Imene Ouali, Mohamed Ben Halima, and Ali Wali. 2022. Text detection and recognition using augmented reality and deep learning. In *Advanced Information Networking and Applications: Proceedings of the 36th International Conference on Advanced Information Networking and Applications (AINA-2022), Volume 1*. Springer, 13–23.
 - [80] Karan Patel and Bhavesh Parmar. 2020. Assistive device using computer vision and image processing for visually impaired; review and current status. *Disability and Rehabilitation: Assistive Technology* (2020), 1–8.
 - [81] Dariusz Plikynas, Arūnas Žvironas, Andrius Budrionis, and Marius Gudauskis. 2020. Indoor navigation systems for visually impaired persons: Mapping the features of existing technologies to user needs. *Sensors* 20, 3 (2020), 636.
 - [82] R. Prabha, M. Razmah, G. Saritha, RM Asha, Senthil G. A., and R. Gayathiri. 2022. Vovoice - Reading Assistant for the Blind using OCR and TTS. In *2022 International Conference on Computer Communication and Informatics (ICCCI)*. 01–07. <https://doi.org/10.1109/ICCCI54379.2022.9740877>
 - [83] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. "Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3173574.3174033>
 - [84] Tuhina Priya, Kotla Sai Sravya, and S Umamaheswari. 2020. Machine-learning-based device for visually impaired person. In *Artificial Intelligence and Evolutionary Computations in Engineering Systems*. Springer, 79–88.
 - [85] Samuel Reinders, Matthew Butler, and Kim Marriott. 2020. "Hey Modell" – Natural User Interactions and Agency in Accessible Interactive 3D Models. 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, 1–13. <https://doi.org/10.1145/3313831.3376145>
 - [86] Frode Eika Sandnes. 2016. What do low-vision users really want from smart glasses? Faces, text and perhaps no glasses at all. In *Computers Helping People with Special Needs: 15th International Conference, ICCHP 2016, Linz, Austria, July 13–15, 2016, Proceedings, Part I* 15. Springer, 187–194.
 - [87] Shailesh Kumar Shivakumar and Shailesh Kumar Shivakumar. 2018. Why Digital Projects Fail? *Complete Guide to Digital Project Management: From Pre-Sales to Post-Production* (2018), 249–265.
 - [88] Walter CSS Simões, Guido S Machado, André MA Sales, Mateus M de Lucena, Nasser Jazdi, and Vicente F de Lucena Jr. 2020. A review of technologies and techniques for indoor navigation systems for the visually impaired. *Sensors* 20, 14 (2020), 3935.
 - [89] M Sreeraj, Jestin Joy, Alphonsa Kuriakose, MB Bhameesh, Anoop K Babu, and Merin Kunjumon. 2020. VIZIYON: Assistive handheld device for visually challenged. *Procedia Computer Science* 171 (2020), 2486–2492.
 - [90] Marc Steen, Menno Manschot, and Nicole De Koning. 2011. Benefits of co-design in service design projects. *International Journal of Design* 5, 2 (2011).
 - [91] Ricardo Tachiquin, Ramiro Velázquez, Carolina Del-Valle-Soto, Carlos A Gutiérrez, Miguel Carrasco, Roberto De Fazio, Andrés Trujillo-León, Paolo Visconti, and Fernando Vidal-Verdú. 2021. Wearable urban mobility assistive device for visually impaired pedestrians using a smartphone and a tactile-foot interface. *Sensors* 21, 16 (2021), 5274.
 - [92] Ruxandra Tapu, Bogdan Mocanu, and Titus Zaharia. 2020. Wearable assistive devices for visually impaired: A state of the art survey. *Pattern Recognition Letters* 137 (2020), 37–52.
 - [93] Mohammad Moeen Valipoor and Angélica de Antonio. 2022. Recent trends in computer vision-driven scene understanding for VI/blind users: a systematic mapping. *Universal Access in the Information Society* (2022), 1–23.
 - [94] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
 - [95] World Health Organization (WHO). 2023. Vision impairment and blindness. <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>
 - [96] Saumya Yadav, Rakesh Chandra Joshi, Malay Kishore Dutta, Martin Kiac, and Pavel Sikora. 2020. Fusion of object recognition and obstacle detection approach for assisting visually challenged person. In *2020 43rd International Conference on Telecommunications and Signal Processing (TSP)*. IEEE, 537–540.
 - [97] Yutaro Yamanaka, Seita Kayukawa, Hironobu Takagi, Yuichi Nagaoka, Yoshimune Hiratsuka, and Satoshi Kurihara. 2022. One-Shot Wayfinding Method for Blind People via OCR and Arrow Analysis with a 360-Degree Smartphone Camera. In *Mobile and Ubiquitous Systems: Computing, Networking and Services: 18th EAI International Conference, MobiQuitous 2021, Virtual Event, November 8–11, 2021, Proceedings*. Springer, 150–168.
 - [98] Ciyuan Yang, Shuchang Xu, Tianyu Yu, Guanhong Liu, Chun Yu, and Yuanchun Shi. 2021. LightGuide: Directing Visually Impaired People along a Path Using Light Cues. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–27.
 - [99] Hanlu Ye, Meethu Malu, Uran Oh, and Leah Findlater. 2014. Current and Future Mobile and Wearable Device Use by People with Visual Impairments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (CHI '14). Association for Computing Machinery, New York, NY, USA, 3123–3132. <https://doi.org/10.1145/2556288.2557085>
 - [100] Analyn N Yumang, Dave Emilson S Banguilan, and Clark Kent S Veneracion. 2021. Raspberry Pi based food recognition for visually impaired using YOLO algorithm. In *2021 5th International Conference on Communication and Information Systems (ICCIS)*. IEEE, 165–169.
 - [101] Yuhang Zhao, Elizabeth Kupferstein, Brenda Veronica Castro, Steven Feiner, and Shiri Azenkot. 2019. Designing AR visualizations to facilitate stair navigation for people with low vision. In *Proceedings of the 32nd annual ACM symposium on user interface software and technology*. 387–402.
 - [102] Yuhang Zhao, Elizabeth Kupferstein, Hathaithorn 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*. 1–14.
 - [103] Yuhang Zhao, Sarit Szpiro, and Shiri Azenkot. 2015. Foresee: A customizable head-mounted vision enhancement system for people with low vision. In *Proceedings of the 17th international ACM SIGACCESS conference on computers & accessibility*. 239–249.
 - [104] Yuhang Zhao, Sarit Szpiro, Jonathan Knighten, and Shiri Azenkot. 2016. CueSee: exploring visual cues for people with low vision to facilitate a visual search task. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 73–84.

A TECH ADOPTION GROUPS

Table 3: Tech Adoption Groups as classified by [22]

Adoption Group	Description
Innovator (I)	A person who builds or develops smart assistive technology
Early Adopter (EA)	A person who buys new smart assistive technology as soon as its available to public
Early Majority (EM)	A person who buys new smart assistive technology if its becoming popular with everyone
Late Majority (LM)	A person who waits until majority leaves feedback and reviews on the device before buying
Laggards (L)	A person who is very conservative and does not like to change their current tools

B DETAILS OF PARTICIPANTS

Table 4: Details of the participants

	Group	Age	Level of Vision	Age of Onset	Tech Adoption
P1	1	55 to 64	LV	Birth	LM
P2	2	55 to 64	LV	Birth	EM
P3	2	18 to 24	LV	Birth	LM
P4	3	45 to 54	LV	26	L
P5	4	55 to 64	LV	Birth	EA
P6	4	45 to 54	LV	17	LM
P7	1	25 to 34	B	3	LM
P8	1	25 to 34	B	Birth	LM
P9	1	65 or over	B	Birth	LM
P10	1	55 to 64	B	9	EM
P11	2	35 to 44	B	16	EA
P12	2	45 to 54	B	38	EM
P13	2	45 to 54	B	Birth	LM
P14	3	35 to 44	B	Birth	EA
P15	3	65 or over	B	46	LM
P16	3	35 to 44	B	Birth	LM
P17	3	55 to 64	B	Birth	EM
P18	4	45 to 54	B	Birth	EA
P19	4	55 to 64	B	Birth	EM
P20	4	45 to 54	B	Birth	EA
P21	4	65 or over	LV	Birth	LM
P22	3	55 to 64	LV	12-14	LM
P23	2	65 or over	LV	64	LM
P24	1	35 to 44	LV	15	EM

C TASKS

Table 5: This table gives the tasks extracted from the review process and their grouping into higher-level categories.

Assistive products for personal mobility
General Navigation
Obstacle Detection - <i>Identify obstacles in the path to navigate safely</i>
Safe Path Detection - <i>Detect and recommend safe paths for navigation</i>
Staircase Detection - <i>Recognise staircases for navigation</i>
Moving Obstacle Detection - <i>Identify dynamic obstacles like gates, revolving doors</i>
Building Door Detection - <i>Identify doors on buildings for entry points recognition</i>
Aerial Obstacle Detection - <i>Detect obstacles that are head height</i>
Queue Detection & Guidance - <i>Detect queues and provide assistance to wait in line</i>
Arrow Detection - <i>Detect and interpret directional arrows for guidance</i>
Safe Navigation Space Detection - <i>Identify safe areas for navigation and waiting</i>
Indoor Navigation
Sign Detection - <i>Identify signs and signage within indoor environments</i>
Room Detection - <i>Recognise different rooms or areas in a building</i>
Shop Isle Guidance - <i>Provide navigation assistance within shop aisles</i>
Outdoor Navigation
Zebra Crossing Identification - <i>Identify zebra crossings on roads or streets</i>
Pedestrian Detection - <i>Identify pedestrians when walking</i>
Pedestrian Signal Detection - <i>Detect pedestrian crossing signal</i>
Bus Stop and Route Identification - <i>Recognise bus stops and their routes</i>
Sidewalk Detection - <i>Detect and locate sidewalks in urban scenes</i>
Terrain Detection - <i>Recognise and categorise different types of terrains</i>
Road Sign Detection - <i>Recognise and locate road signs</i>
Iconic Places Recognition - <i>Recognise famous landmarks and iconic locations</i>
Braille block detection - <i>Identify Braille blocks or access tiles on the ground</i>
Bus Detection - <i>Detect approaching or stopped buses</i>
Animal Waste or Waste Detection - <i>Locate and identify animal waste or litter</i>
Bus Door Detection - <i>Identify bus doors for boarding</i>
Driveway Detection - <i>Detect driveways in residential or commercial areas</i>
Shop Recognition - <i>Identify and name shops or stores</i>
Public Transport Board Detection - <i>Identify and read boards displaying public transport information</i>
Water Body Detection - <i>Detect and identify bodies of water such as lakes, rivers, etc.</i>

Assistive products for communication and information
Human Interaction
Face Detection - <i>Locate faces and identify them</i>
Emotion Detection - <i>Identify emotions of people</i>
Gesture Detection - <i>Recognise hand or body movements pointing, raising hands</i>
Human Pose or Activity Detection - <i>Recognise body poses or actions such as sleeping, sitting</i>
Social Distance Detection - <i>Measure and help maintain safe/comfortable distances from people</i>
Family Member Identification - <i>Locate and Identify family members</i>
Face Detection for Selfies - <i>Find faces specifically in selfie-style images to help frame better</i>
Age and Gender Recognition - <i>Determine age and gender of people</i>
Count of people in a room - <i>Estimate the number of individuals present</i>
Eye Contact Detection - <i>Detect whether eye contact is made during conversation</i>
Hand Detection - <i>Detect and track human hands to pass and receive objects</i>
Social Helper Detection - <i>Identify individuals offering assistance such as policeman</i>
Textual and Graphical Information
Text Recognition - <i>Detect and read text</i>
Barcode & QR Code Reader - <i>Scan and decode barcodes and QR codes</i>
Braille Recognition - <i>Convert and read Braille text</i>
Color Detection - <i>Identify and detect colors</i>
Fingertip Detection - <i>Detect and read text pointed to the finger tip of the user</i>
Fill paper forms - <i>Assist with filling paper forms</i>
Reading Time - <i>Detecting clock faces to read time</i>
Environmental Information
Virtual Assistant - <i>Allows user to ask questions about their environment</i>
Image Captioning - <i>Generate description of what the user is facing</i>
Scene Query / VQA - <i>Allows the user to ask any query and maintain a conversation about their surrounding</i>
Scene Recognition - <i>Identify and categorise the type of environment such as a hospital room or a parking lot</i>
Point to Recognise - <i>Identify objects or elements pointed at by users</i>
Scene Enhancement - <i>Enhances contrast by distinguishing objects with different colours</i>
Day-Night Detection - <i>Distinguish whether its day time or night time</i>
Overlays - <i>Superimpose digital information or image enhancement onto real-world scenes</i>
Scene Simplification - <i>Reduce visual clutter and detail in the user's field of view to enhance comprehension</i>

Assistive products for handling objects and devices
General Objects
Object Detection and Localization - <i>Identify and locate various objects</i>
Object Manipulation Aid - <i>Assist users in interacting with and manipulating objects</i>
Specific Objects
Currency Detection - <i>Recognise and differentiate currencies for transactions</i>
Personal Object Recognition - <i>Identify personal belongings for organisation and tracking</i>
Food Recognition - <i>Identify and classify different types of food items</i>
Cloth Pattern Recognition - <i>Detect and categorise patterns on clothing items</i>
Appliance Recognition - <i>Identify household appliances and their controls</i>
Pen Tip Recognition - <i>Detect and locate writing pen tips</i>
Document Type Detection - <i>Identify document types such as passports, credit cards, insurance cards</i>

Assistive products for personal care and protection
Fall Detection - <i>Detect and alert for potential fall incidents in real-time</i>
Fire Detection - <i>Detect and alert for fire incidents</i>
Thermal Hazard Detection - <i>Identify items with hazardous temperature levels like a hot kettle</i>

Assistive products for cultural and sports activities
Artwork Accessibility - <i>Detect and provide descriptions of artworks</i>
Museum Accessibility - <i>Assist navigation by provide guidance and descriptions of museums artefacts</i>
Physical Games Accessibility - <i>Detect and provide descriptions of physical games</i>

Assistive products for personal medical treatment
Drug Pill Detection - <i>Recognise and describe medications</i>