PROCEEDINGS of the
24th International Congress on Acoustics

October 24 to 28, 2022 in Gyeongju, Korea

Acoustic touch: An auditory sensing paradigm to support close reaching for people who are blind

Craig Jin(1), Julee-anne Bell(5), Lil Deverell(5), Felicity Gates(5), Ibai Gorodo(5), Shayikh Hossain(1), Chin-Teng Lin(2), Marx Melencio(5), Minh Nguyen(4), Vincent Nguyen(3), Avinash Singh(2), Howe Zhu(2)

(1)School of Electrical and Information Engineering, The University of Sydney, Australia,
{craig.jin,shayikh.hossain}@sydney.du.au
(2)Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia,
{chin-teng.lin,vincent.nguyen,avinash.singh,howe.zhu}@uts.edu.au
(3)Graduate School of Health, University of Technology, Sydney, Australia, vincent.nguyen@uts.edu.au
(4)Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia,
minh.t.nguyen-4@student.uts.edu.au
(5)Research and Development, ARIA Research Pty Ltd, Australia, {julee-anne,lil,felicity,ibai,marx}@ariaresearch.com.au

ABSTRACT

This work explores an auditory sensory augmentation paradigm we call acoustic touch, to assist people who are blind with reaching for close objects. The sensory augmentation system is constructed based on the Nreal augmented-reality glasses using a custom application running on an android phone. The system recognizes and localizes objects visually using cameras in the glasses, then renders objects as sound within a limited field-of-view, so we shall refer to the glasses as a foveated audio device. The repetition of the sound varies depending on the location of the object within the field of view of the foveated audio device. Psychophysical tests of the spatial perception of multiple objects are conducted comparing the acoustic touch paradigm with two other conditions: (1) a verbal clock face description of object locations and (2) a sequential audio presentation of the objects using Bluetooth speakers located with the objects. We report on the results of the psychophysical study with blind and blindfolded sighted participants.

Keywords: Auditory Sensory Augmentation, Acoustic Fovea, Acoustic Touch

1 INTRODUCTION

The development of assistive technologies for people with blindness or low vision has a long history (for review see [1]) with considerable work addressing spatial understanding and mobility that includes orientation and navigation, obstacle avoidance, hazard minimization, and information & signs. Broadly, sensory substitution refers to one sensory modality feeding information to another, such as vision-to-hearing. However, this description provides a poor understanding of the possibilities given modern technologies. These days we can integrate information from multiple sensors simultaneously (e.g., camera vision, motion sensors, etc.) and process the data using powerful machine learning algorithms to implement novel sensory augmentation paradigms with specific objectives. Such sensory augmentation paradigms can introduce new information streams that may add value and richness to the sensory experience. Significantly, we subscribe to the relatively recent view [2, 3] that it is better to view sensory substitution as an acquired cognitive extension rather than as a modal perceptual experience such as hearing, vision, and touch. In this view, cognitive aspects come to the fore, including information processing and flow, learning and training, and neural plasticity.

In this work, we consider a potential assistive technology for people with blindness or low vision which is based on auditory sensory augmentation via smart glasses. The application of smart glasses with augmented reality for assistive technologies is a rapidly growing field [4] and there are many augmented reality glasses on the market [5]. For this work, we develop and test a system for close reaching based on the Nreal Light glasses (Nreal, China). The basic principle of operation is to sonify objects within a defined field of view of the glasses. The sonification of information has a long history (for review see [6]). When considering assistive navigation, there are many types of "beacon" sounds such as auditory icons, earcons, spearcons, and morphocons (for review refer to [7]). In this work, we chose to use auditory icons - four unique sounds - to represent four objects commonly found on a table: book, bottle, bowl, and cup. The framework for the study is to have a blind or blindfolded participant seated at a table wearing augmented reality smart glasses that sonify four objects as auditory icons based on machine vision and object recognition. The perceptual task relates to the use of head scanning to develop a spatial map of the objects within reach on the table.

While the framework for the psychophysical study is straightforward, we highlight some of the more subtle issues. To begin, normal human hearing is panoramic, i.e., we hear sound from all directions simultaneously and use spatial hearing [8, 9] to distinguish source location. In this sense, it is quite unnatural to restrict the hearing of sound to a particular field of view. Nevertheless, this of course, is exactly what most auditory sensory augmentation systems based on the camera vision of smart glasses will do because the cameras generally do not support 360 degrees of vision. For convenience, we refer to the field of view as an "acoustic fovea" and think of the device as a foveated audio device (FAD) providing a window on which to hear the world. Please note that this is quite different to the acoustic fovea of high frequency resolution associated with bats and echolocation or the acoustic fovea associated with attention and auditory scene analysis. The introduction of an acoustic fovea via smart glasses immediately raises the issue of head scanning to explore the world. As one turns the head, one encounters new objects as auditory icons. We refer to the combined action of head scanning and the sonification of objects as auditory icons as an "acoustic touch". We make this reference because as one turns the head and the smart glasses activate sound from an object - it is as if one is "touching" the object and acoustically activating it. It is worth highlighting that head scanning may not be common for people with blindness. The focus of this study then is to characterise the accuracy and timing of the acoustic touch with respect to a spatial mapping and reaching task.

There are numerous existing systems and/or services with object recognition objectives similar to the close-reach system in our study. A number of these, however, do not directly consider close-reaching. For example, there is the Orcam MyEye 2.0 (Orcam, Israel); several mobile phone recognition apps: TapTapSee, Lookout, Seeing AI, Supersense; and human visual interpretation services available as an app: Be My Eyes, Aira. There are also numerous smart glasses projects with object recognition objectives such as Envision glasses [**?**], the CARA project using a Microsoft Hololens [10] and custom systems created using, e.g., a Raspberry Pi [11]. As well, there have been numerous navigation aids proposed [10, 12, 13, 14], for a review see [15]. In this light, the contributions of this paper are: (1) to describe a real-time system for the exploration of auditory sensory augmentation using the Nreal light glasses; (2) to describe a measurement framework for the study including motion capture and physiological sensors; and (3) to provide some new results on a spatial mapping and close reaching task. Section 2 describes the methods, Section 3 describes the results and we then provide a discussion in Section 4 and conclude.

2 METHODS

We describe the methods for the current study in three parts: (1) the smart glasses system; (2) the psychophysical experiment; and (3) the psychophysical measurement system. We begin with the smart glasses system.

2.1 Smart Glasses System

The main component of the smart glasses system used in this study is a pair of Nreal Light glasses (Nreal, China) consisting of cameras, an inertial sensor, and a pair of stereo speakers. The sensor data coming from the sensors in the glasses are used to capture the environment around the user while the speakers in the glasses are used to represent the captured environment in sound. Since the Nreal Light glasses have no processing capability, the glasses are connected via USB to an Oppo Find X3 Pro Android smartphone (Oppo, China). The data are transmitted from the glasses to perform all the data processing on the phone using a custom app (refer to Fig. 2). This processing is divided into two parts: a perception and a sonification system. The main responsibility of the perception system is to extract information from the sensors in order to populate a virtual world that resembles the actual location of objects in the real world.

Object recognition is implemented using a YOLOv5 object detector to capture the position of objects in the images from the glasses' RGB camera. The object detector used in this study was trained by the original authors on the COCO dataset [16] to be able to capture 80 different types of objects. Out of those 80 objects, the perception system filters out the results to only get the results of the following four object classes: book, bottle, bowl, and cup. For the purposes of this study, these objects were selected based on the types of objects that are commonly found on a table. The perception system then combines the detected objects with the estimated stereo depth and head tracking to obtain the 3D position of the detected objects in a fixed coordinate system. Finally, an object tracker uses the 3D position of these objects to identify and track these objects over time.

The phone app sonification system takes the extracted information regarding the detected objects and generates a 3D spatial auditory icon that matches the type of object being sonified and the location of the object. The Unity spatial audio engine is used. For this purpose, the captured objects are first populated into a virtual world in Unity (Unity Technologies, USA) to generate a game-like environment where the user is interacting with these objects by moving their head to trigger the objects to generate a sound. While the auditory icons are sonified using the Unity spatial audio engine, we do not consider this a spatial audio application because the sounds are restricted by the field of view. It is the interaction between the field of view of the FAD and the head scanning of the participant that provides spatial information. The field of view established by the phone app is adjustable and controls the angular range in which objects are sonified. Note that the field of view can be adjusted in software so that the "acoustic fovea" is larger than the field of view of the cameras. We experimented with the field of view in several face validity tests and settled on a value of 36.6° horizontally and 20° vertically. The repetition period, ΔR, for the auditory display of the auditory icons depends on the camera angle of the object, α, and is given by the following formula:

$$\Delta R = 0.02\left(\left\lfloor \frac{\alpha}{36.6} \right\rfloor + 1\right) , \tag{1}$$

where α is assumed to be in degrees and $\lfloor\cdot\rfloor$ is the floor function rounding down to the nearest integer.

2.2 Psychophysical Experiment

2.2.1 Participants

The recruitment criteria for the blind participants were adults over the age of 18, any gender with blindness or ultra-low vision (<LogMar 2.0), and using a long cane or guide dog to support safe, independent travel skills. The recruitment criteria for the sighted participants were adults over the age of 18, any gender, and comfortable with working under blindfold. We recruited sixteen participants for the psychophysical experiment, nine blind participants (M/F: 6/3; average age: 53.0) and seven sighted participants (M/F: 3/4; average age: 42.8).

2.2.2 Sounds

Four representative auditory icons were chosen for the four table objects as shown in Table 1. The time signals and spectra for the auditory icons are shown in Fig. 1.

Table 1. Auditory Icons

Object	Sound
Book	Turning pages
Bottle	Scraping a glass bottle
Bowl	Lid placed on a bowl
Cup	Cup placed on a wooden table

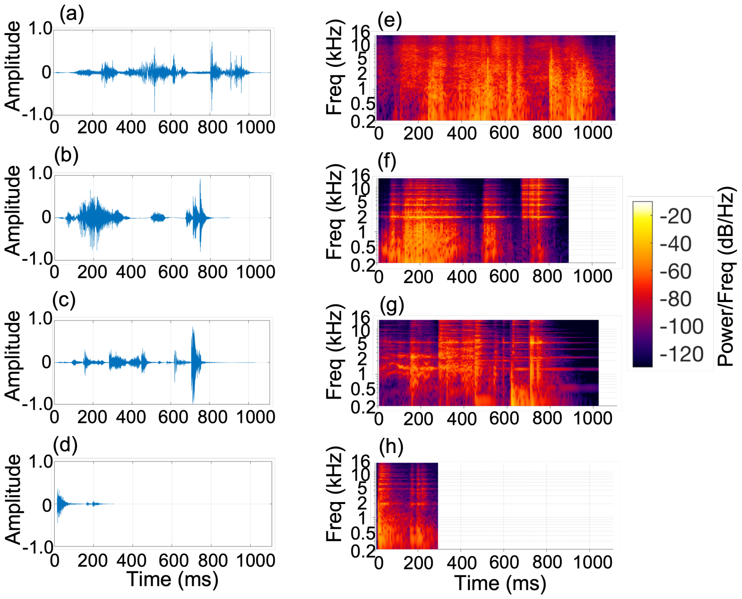

Figure 1. The time signals (a)-(d) and spectra (e)-(h) of the four auditory icons are shown in the following order: book, bottle, bowl, and cup, respectively.

2.2.3 Experimental Conditions and Task

There are three experiment conditions in which the primary psychophysical task is a seated spatial mapping and close-reaching task. In each condition, the participant is provided information regarding the identity of objects on the table and their location. Based on the provided information, the participants are required to create a mental spatial map of the objects on the table so that they may later reach out and touch a specified object. Significantly, no information or feedback is provided during the reaching task itself. The three experiment conditions are described below:

- **Foveated Audio Device:** The Nreal glasses and smartphone app are used with head-scanning for auditory icon sonification.
- **Bluetooth speaker:** A small Bluetooth speaker is placed on the table at the location of each object, coded to play the auditory icon corresponding to its matching object three times. The auditory icons play in a clockwise arc, from left to right.
- **Clockface instructions:** The location and identity of objects are described verbally relative to a clock face, with 9 o'clock on the left of the table, midday ahead, and 3 o'clock on the right of the table, in a clockwise order from left to right.

The number of objects placed on the table for each trial is randomly chosen to be either two or three objects. The order of operation is as follows. An experimenter quietly sets objects on the table; the trial then begins and the participant starts the mental spatial mapping task. When ready, having completed the spatial mapping task, the participant presses a large button with their left hand and this triggers a verbal computer instruction to touch the target object. The participant then stretches out their right hand to touch the target object directly. Note that during the FAD condition, the participant is asked to identify the objects detected on the table before pressing the instruction button and reaching for the target object.

2.2.4 Training

All participants were given a chance to familiarize themselves and train for the psychophysical task. There are a number of activities requiring training: learning auditory icons, becoming comfortable with head-scanning,

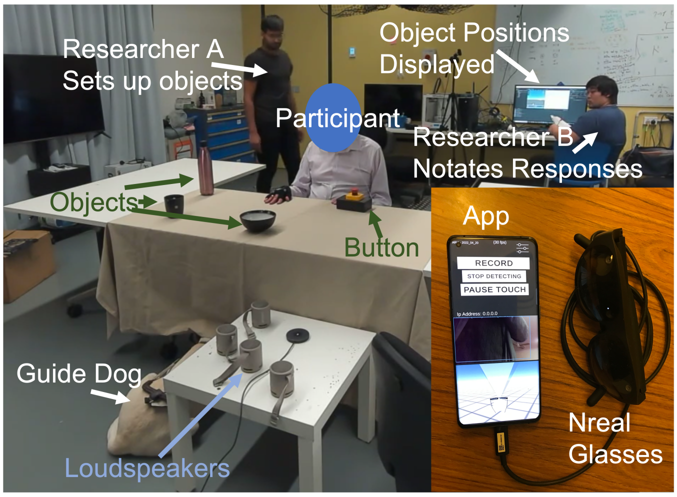

Figure 2. The experimental setup is shown.

building spatial maps based on each of the three experiment conditions, reaching, narrating, and using the tactile push button. We found that the training time varied widely across participants with some participants requiring sessions on two days to complete the experiment, while other participants completed the experiment in one session. Once the participants were confident with the tasks, a dry run of at least ten trials were completed under official experiment conditions to ensure participants were comfortable with the procedures.

2.3 Experimental Measurement System

This experiment used a software experiment manager to instruct the researchers on the trial order, object layout, and resting periods (refer to Fig. 2). The manager system was programmed on the Unity game engine. During each experiment session, the experiment manager software produces a pseudo-randomized trial sequence with the three conditions interweaved in a manner that is balanced across trial blocks of 20 and that minimises trial repetitions such that there is a maximum of three repetitions for a given trial. The current condition of the trial and the object layout on the table were displayed to the researchers as a map on a computer monitor. The researcher or assistant used the displayed trial map to place the objects accordingly. For the Bluetooth speaker condition, the trials required a Bluetooth speaker to be placed in front of each object. The Sony SRSXB13 speakers were used for this experiment. Each speaker was paired to a Bluetooth transmitter which was connected to a channel on the RME Fireface 400 sound card. Four speakers were used in this experiment (one for participant instructions and three for playing object sounds). The experiment manager communicates with the firewire sound card through a multichannel audio playback Unity asset [17].

The participant's performance are determined through observation and the use of sensors. During the experiment, a researcher noted the participant's results on a trial runsheet, recording whether the participant successfully reached the object, if they missed or mixed up the object with another object, and the objects successfully identified during the FAD trials. A streaming server software known as Lab Streaming Layer (LSL) was used to synchronize the timing between the participant input and the trial times. The large tactile button that prompted instructions for the reaching task was connected to the experiment manager computer via a USB dongle receiver. The participant response times were calculated using event times recorded by LSL. Refer to Fig. 2 for a photo of the experimental setup.

The experiment was additionally recorded on video and motion capture. Twelve Flex13 Optitrack motion capture cameras were used to record the positions of the objects, and the participant's head and hand positions. In addition, physiological measures such as heart-rate variability (Zephyr Bioharness) and electrodermal activity (E4 Wristband) were recorded during the experiment. A number of self-assessment questionnaires were also used to measure the participants experience: a questionnaire regarding experience with prior assistive devices, a

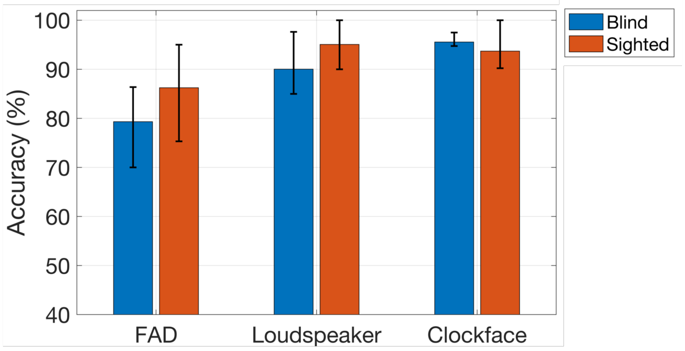

Figure 3. The mean performance accuracy for the reaching tasks are shown for the three experimental conditions and both the blind and sighted participant groups. The coloured bars indicate the mean performance accuracy, while the black bars indicate the inter-quartile range (IQR).

Karolinska sleepiness scale questionnaire, and three iterations of the NASA Task Load Index (TLX) questionnaire. The data from the physiological sensors and the questionnaires are not reported upon here, but will be reported on in future presentations.

3 RESULTS

The experimental results for the spatial mapping and reaching task are now described. To begin, a few participants were often confused by the auditory icons and had seemingly not had sufficient time to learn and distinguish the sounds. As well, a few participants seem to confuse the clock face directions, e.g., mixing 3 o'clock with 9 o'clock. The outlying nature of these results indicated that insufficient training time had been given. We decided to remove these participant's data for a given experiment condition if the participant could not complete the task at above 50% accuracy for all objects in that condition. The application of the above 50% accuracy rule, leads to participant numbers for each experiment as shown in Table 2.

Table 2. Participant Numbers for Each Experiment Condition

Experiment Condition	Blind Participants	Sighted Participants
Foveated Audio Device	9	3
Bluetooth speaker	8	6
Clockface instructions	8	7

In Fig. 3 a bar plot shows the performance accuracy for the reaching task for the three experimental conditions for both the blind and sighted participant groups: (1) the FAD with the acoustic touch (B,S: 79.3 ± 10.1, 86.3 ± 15.1); (2) the Bluetooth speaker condition (B,S: 90.0 ± 10.0, 95.1 ± 4.5); (3) and the Clockface description (B,S: 95.6 ± 3.2, 93.7 ± 7.4). A Kruskal-Wallis test comparing the blind and sighted participant groups for each experimental condition shows no statistically significant differences ($H(1) < 0.97$, $p > 0.32$). Both participant groups demonstrate reasonable performance accuracy (> 79%) in the FAD condition, nevertheless, the performance accuracy is less than that for the other two conditions. We discuss the difference in performance accuracy in more detail in Section 4. For the blind cohort, a Kruskal-Wallis test comparing the performance accuracy for the FAD condition with each of the other two conditions separately, showed statistically significant differences (H(1) > 4.33, p < 0.037) with the following relative success rate differences (Cliff's delta), Bluetooth speaker: $\delta = 0.63$, Clockface: $\delta = 0.82$.

Consider now the individual objects and the accuracy of the spatial mapping and reaching performance

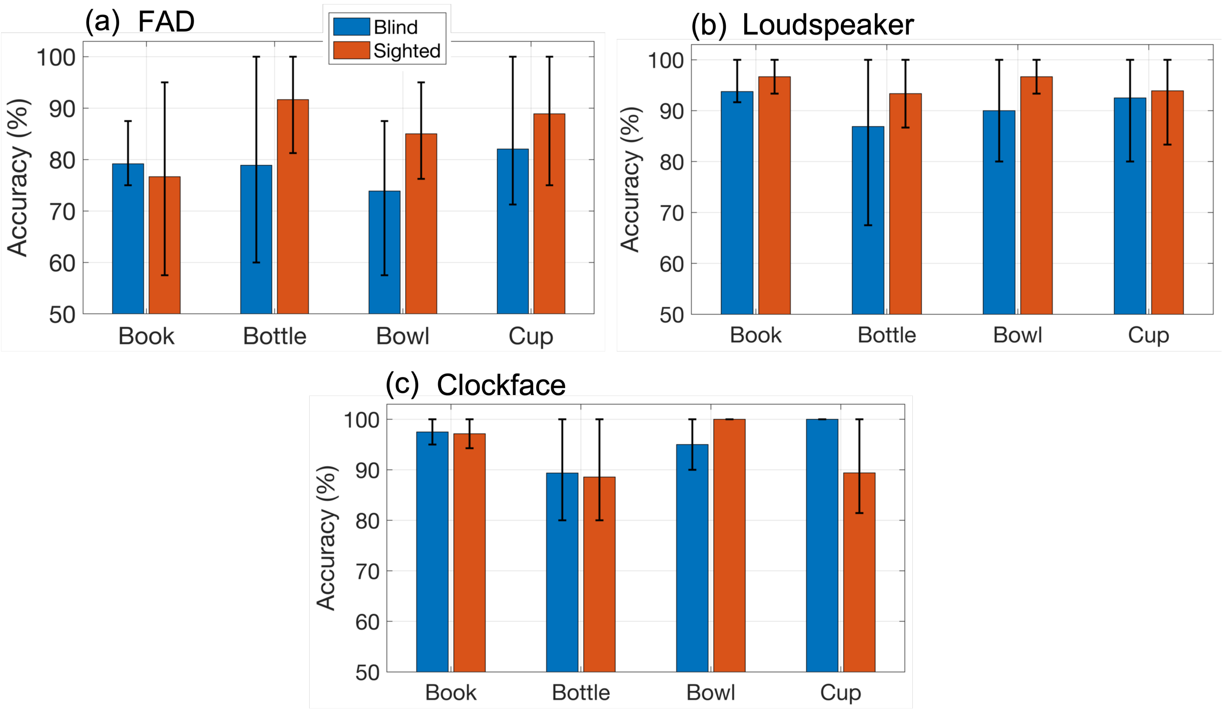

Figure 4. The accuracy of reaching is shown for each of the four objects (book, bottle, bowl, cup) for all three experimental conditions: (a) FAD, (b) Loudspeaker, (c) Clockface, and both the blind and sighted participant groups. The coloured bars indicate the mean performance accuracy, while the black bars indicate the inter-quartile range (IQR).

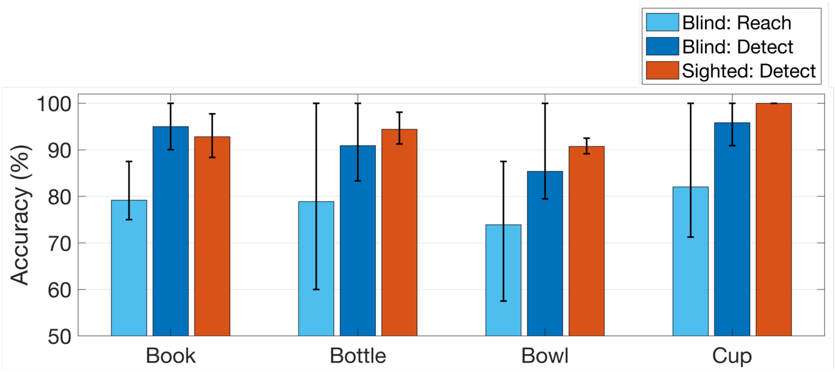

Figure 5. The performance accuracy for the detection task for the four objects is shown for the FAD condition for both the blind and sighted participant groups. The coloured bars indicate the mean performance accuracy, while the black bars indicate the inter-quartile range. For comparison, the performance data for the reaching task is also shown for the blind participant group.

as shown in Fig. 4. We see that performance is similar across objects and a Kruskal-Wallis test comparing performance differences across the four objects shows no significant differences for the blind participants ($H(3) = 0.72$, $p = 0.87$) or sighted participants ($H(3) = 0.07$, $p = 0.80$).

Spatial mapping and reaching is a more complicated perceptual task than detecting and reporting an object. In this regard, Fig. 5 shows the performance accuracy for the simpler task of detecting and reporting each of the four objects in the FAD condition. For reference, we also show the reaching task performance for the blind participants in Fig. 5. Comparisons between the detection and reaching tasks indicate that the blind participants show the following increased accuracy of performance for object detection compared with object reaching, book: 15.8%, bottle: 12.0%, bowl: 11.5%, cup: 13.8%. Combining the data across objects and applying a Kruskal-Wallis test comparing the detection and reaching tasks indicate these differences are statistically significant ($H(1) = 15.49$, $p < 0.0001$).

The scan time for the FAD represents the exploratory time during which the participants used head scanning to create a mental spatial representation of the objects located on the table. The average scan time across all of the FAD trials for each participant is shown in Fig. 6. We see a wide range of average scan times varying from a minimum of 22.2 s to a maximum of 94.2 s. The average of the mean trial scan times across the participants is 45.0 s with a standard deviation of 20.3 s. It is useful to put these scan times into context. We can compare these scan times with the time taken to both listen and react to the stimuli presented during the Bluetooth speaker and Clockface conditions, as shown in Table 3. We see that the average scan time for the FAD condition is approximately 2.25 times the listen-and-react time for the Bluetooth speaker condition and 2.59 times the listen-and-react time for the Clockface condition. A critical issue related to the average scan times is whether or not these times are correlated with the performance accuracy of the spatial mapping and reaching task. To address this matter, we show in Fig. 7 a scatter plot of the performance accuracy for the reaching task versus the average trial scan times for the various participants. We see that there is no general correlation between head-scanning time and performance accuracy in reaching ($R^2 = 0.002$). Further, three subjects demonstrate a 95% performance accuracy for reaching with head-scanning times varying from 33.6 s to 42.0 s, which are much less that the two longest head-scanning times of 63.2 s and 94.2 s.

Table 3. Head-scanning and listen-react times

Condition	Activity	Mean Time (s)	Standard Deviation (s)
FAD	Head-Scanning	45.0	20.3
Bluetooth speaker	Listen and React	20.0	0.8
Clockface	Listen and React	17.4	0.6

4 DISCUSSION

The close-reaching task in this study is a mental spatial mapping task that requires spatial memory. In other words, no auditory information is provided to the participant during the actual reaching task. The provision of auditory information was deliberately separated from reaching during the psychophysical study to better explore the element of spatial mapping. This is different to many long-range navigational studies [13, 12, 14] which provide continuous information during the task and therefore the opportunity for sensory confirmation and integration. In this sense, the current study is most similar to what is more commonly referred to as near-field assistance [13] and/or spatial memory tasks [10]. The results presented here are consistent with previous studies indicating the benefits of smart glasses for the blind with respect to object localization and identification [18, 10, 19, 20]. For example, in [18] blind participants wearing a smart glasses system were able to successfully perform scene classification tasks based on audio rendering. In [10] blind participants using a Hololen's system were able to successfully perform a spatial memory task in which five objects are placed 30 degrees apart in azimuth at a two meter distance and rendered using spatialized verbal audio cues. In this

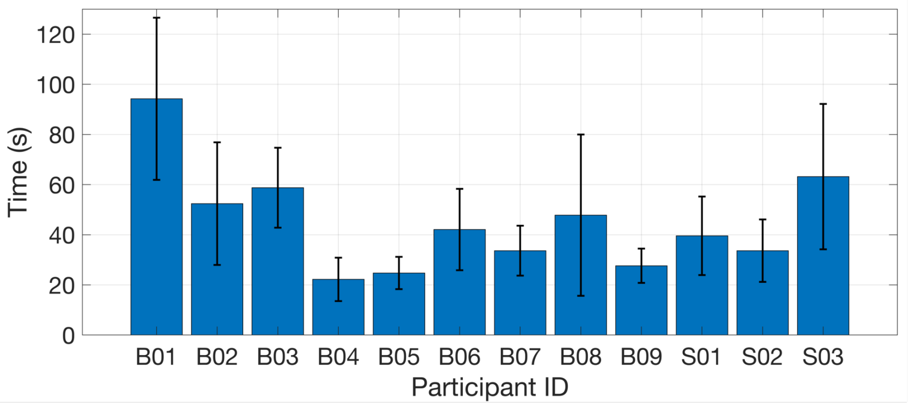

Figure 6. The average scan time is shown using coloured bars for the blind participants (ID starts with a 'B') and sighted participants (ID starts with an 'S') with the black bars indicating ± one standard deviation.

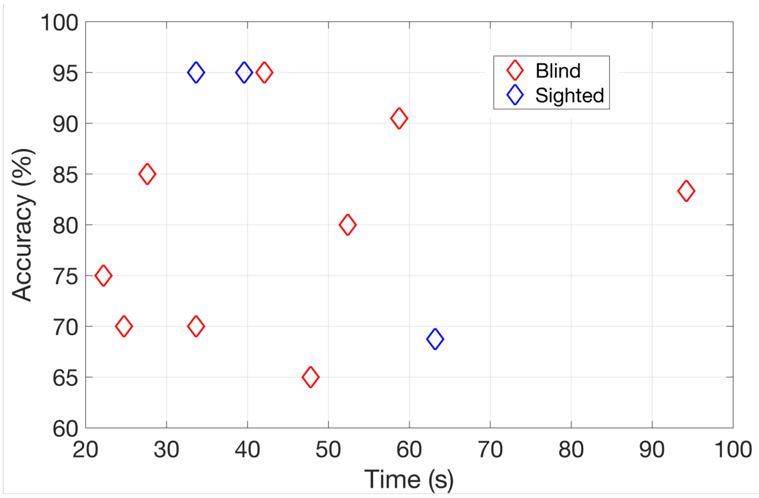

Figure 7. A scatter plot is shown of the performance accuracy for reaching in the FAD condition versus the mean trial head-scanning time. The plot indicates no general correlation between reaching accuracy and head-scanning time ($R^2 = 0.002$).

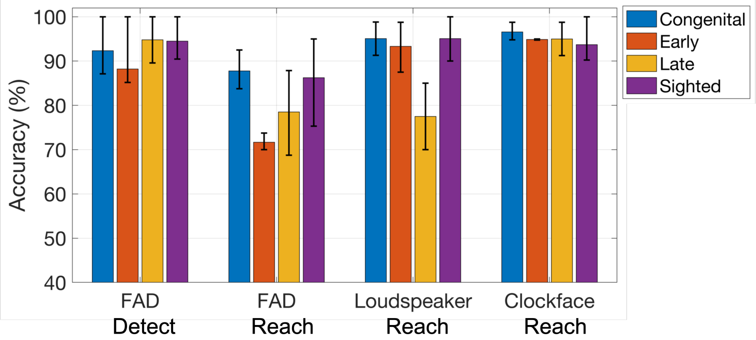

Figure 8. The performance accuracy for detection and reaching tasks are shown for the blind participants based on the onset of blindness: congenital, early (in childhood), and late (as an adult). Data for the sighted participants are shown for reference.

test, participants had to orient their head toward the target object. As well, in [19] blind participants were able to use a smart glasses system to recognize currency.

In terms of studying spatial memory and cognition, particularly in the context of models of working memory [21], tasks are often divided into active tasks in which one has to actively work with the information and passive tasks in which information is simply recalled [22]. It is generally agreed that it is the active tasks that show differentiated performance between blind and sighted participants [23, 22]. In these terms, the close-reaching task presented here has an active component in which reaching must be performed. Nonetheless, the task is much simpler than some of the spatial memory tasks presented previously such as matching the locations of paired spatial sounds [22], or an audio version of the Kohs Block Design test [24] or an audio version of the Corsi-Block test [23]. The results of this study did not show statistically significant differences between blind and sighted participants. This possibly reflects the simpler nature of the active task presented in this study.

Consider now the increasing, but conditional, agreement that blindness impacts allocentric spatial tasks, particularly large-scale spatial tasks [23, 25, 26, 27] (for review see [25]). As the close-reaching task presented here is an egocentric task, we would not expect differences based on onset of blindness: congenital, early onset in childhood, and late onset as an adult. Nonetheless, this does not quite seem to be the case. It turns out that we had equal numbers (three each) of congenital, early and late onset participants. The performance results for the detection and close-reaching tasks are separated based on onset of blindness in Fig. 8. As shown, there seem to be some performance differences appearing with regard to the close-reaching task for the FAD and Bluetooth speaker conditions based on onset of blindness that are not apparent in the detection task or the Clockface condition. Of course, the very small numbers of subjects limits any conclusions. Intriguingly, it is suggested in [28] that self-motion, such as with the head scanning in the FADs condition, can assist the blind with spatial tasks, even on allocentric tasks.

5 CONCLUSIONS

This work investigates an auditory sensory augmentation paradigm, acoustic touch, to assist people who are blind with reaching for close objects. It is based on the use of smart glasses combined with a limited field of view, i.e., an acoustic fovea, and head scanning. Both blind and blind-folded sighted participants are able to perform the task reasonably well demonstrating the utility of the sensory augmentation paradigm. The primary task is an active spatial memory task which demonstrates no statistically significant differences between the blind and sighted participants which is consistent with the fairly uncomplicated nature of the reaching task. Further analysis of the dataset will explore the motion capture and physiological data. We will continue to develop new auditory sensory augmentation paradigms based on smart glasses and machine vision and learning.

ACKNOWLEDGEMENTS

We would like to thank both ARIA Research Pty Ltd and the Australian government for their funding support via a CRC Projects Round 11 grant. We would particularly like to thank Robert Yearsley, Mark Harrison, Glenn Dickins, and Daniel Kish for their support and comments.

REFERENCES

[1] Hersh MA, Johnson MA, editors. Assistive Technology for Visually Impaired and Blind People. London: Springer London; 2008. Available from: `http://link.springer.com/10.1007/978-1-84628-867-8`.

[2] Deroy O, Auvray M. Reading the World through the Skin and Ears: A New Perspective on Sensory Substitution. Frontiers in Psychology. 2012;3. Available from: `http://journal.frontiersin.org/article/10.3389/fpsyg.2012.00457/abstract`.

[3] Kirsch LP, Job X, Auvray M. Mixing up the Senses: Sensory Substitution Is Not a Form of Artificially Induced Synaesthesia. Multisensory Research. 2020 Jul;34(3):297–322. Available from: `https://brill.com/view/journals/msr/34/3/article-p297_5.xml`.

[4] Zuidhof N, Ben Allouch S, Peters O, Verbeek PP. Defining Smart Glasses: A Rapid Review of State-of-the-Art Perspectives and Future Challenges From a Social Sciences' Perspective. Augmented Human Research. 2021 Dec;6(1):15. Available from: https://link.springer.com/10.1007/s41133-021-00053-3.

[5] Stream C. 16 Augmented Reality Glasses of 2021 (with Features); 2022. Available from: https://circuitstream.com/blog/16-augmented-reality-glasses-of-2021-with-features-breakdowns/.

[6] Worrall D. Sonification Design: From Data to Intelligible Soundfields. Human–Computer Interaction Series. Cham: Springer International Publishing; 2019. Available from: http://link.springer.com/10.1007/978-3-030-01497-1.

[7] Parseihian GT, Katz BFG. Morphocons: A New Sonification Concept Based on Morphological Earcons. J Audio Eng Soc. 2012;60(6):10.

[8] Carlile S. Virtual Auditory Space. Berlin, Heidelberg: Springer Berlin / Heidelberg; 2013. OCLC: 1066199040. Available from: https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=5577220.

[9] Blauert J. Spatial Hearing: The Psychophysics of Human Sound Localization. The MIT Press; 1996. Available from: https://direct.mit.edu/books/book/4885/Spatial-HearingThe-Psychophysics-of-Human-Sound.

[10] Liu Y, Stiles NR, Meister M. Augmented reality powers a cognitive assistant for the blind. eLife. 2018 Nov;7:e37841. Available from: https://elifesciences.org/articles/37841.

[11] Sharma S, Kalra N, Gupta L, Varma N, Agrawal S, Verma V. VASE: Smart glasses for the visually impaired. Journal of Ambient Intelligence and Smart Environments. 2022 May;14(3):213–226. Available from: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/AIS-210491.

[12] Bujacz M, Skulimowski P, Strumiłło P. Naviton—A Prototype Mobility Aid for Auditory Presentation of Three-Dimensional Scenes to the Visually Impaired. J Audio Eng Soc. 2012;60(9):13.

[13] Katz BFG, Kammoun S, Parseihian G, Gutierrez O, Brilhault A, Auvray M, et al. NAVIG: augmented reality guidance system for the visually impaired: Combining object localization, GNSS, and spatial audio. Virtual Reality. 2012 Nov;16(4):253–269. Available from: http://link.springer.com/10.1007/s10055-012-0213-6.

[14] Geronazzo M, Bedin A, Brayda L, Campus C, Avanzini F. Interactive spatial sonification for non-visual exploration of virtual maps. International Journal of Human-Computer Studies. 2016 Jan;85:4–15. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1071581915001287.

[15] Hersh MA, Johnson MA. Mobility: An Overview. In: Hersh MA, Johnson MA, editors. Assistive Technology for Visually Impaired and Blind People. London: Springer London; 2008. p. 167–208. Available from: http://link.springer.com/10.1007/978-1-84628-867-8_5.

[16] Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft COCO: Common Objects in Context. In: ECCV. European Conference on Computer Vision; 2014. Available from: https://www.microsoft.com/en-us/research/publication/microsoft-coco-common-objects-in-context/.

[17] Low-latency multichannel audio: Audio Sound FX. Unity Asset Store;. Available from: https://assetstore.unity.com/packages/audio/sound-fx/low-latency-multichannel-audio-147091.

[18] Ribeiro F, Florencio D, Chou PA, Zhang Z. Auditory augmented reality: Object sonification for the visually impaired. In: 2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP). Banff, AB, Canada: IEEE; 2012. p. 319–324. Available from: http://ieeexplore.ieee.org/document/6343462/.

[19] Parlouar R, Dramas F, Macé MMJ, Jouffrais C. Assistive device for the blind based on object recognition: an application to identify currency bills. In: Proceeding of the eleventh international ACM SIGACCESS conference on Computers and accessibility - ASSETS '09. Pittsburgh, Pennsylvania, USA: ACM Press; 2009. p. 227. Available from: http://portal.acm.org/citation.cfm?doid=1639642.1639688.

[20] Dramas F, Oriola B, Katz BG, Thorpe SJ, Jouffrais C. Designing an assistive device for the blind based on object localization and augmented auditory reality. In: Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility - Assets '08. Halifax, Nova Scotia, Canada: ACM Press; 2008. p. 263. Available from: http://portal.acm.org/citation.cfm?doid=1414471.1414529.

[21] Baddeley A. Working Memory. Science. 1992 Jan;255(5044):556–559. Available from: https://www.science.org/doi/10.1126/science.1736359.

[22] Setti W, Cuturi LF, Cocchi E, Gori M. Spatial Memory and Blindness: The Role of Visual Loss on the Exploration and Memorization of Spatialized Sounds. Frontiers in Psychology. 2022 May;13:784188. Available from: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.784188/full.

[23] Setti W, Cuturi LF, Cocchi E, Gori M. A novel paradigm to study spatial memory skills in blind individuals through the auditory modality. Scientific Reports. 2018 Dec;8(1):13393. Available from: http://www.nature.com/articles/s41598-018-31588-y.

[24] Bertonati G, Tonelli A, Cuturi LF, Setti W, Gori M. Assessment of spatial reasoning in blind individuals using a haptic version of the Kohs Block Design Test. Current Research in Behavioral Sciences. 2020 Nov;1:100004. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2666518220300048.

[25] Pasqualotto A, Proulx MJ. The role of visual experience for the neural basis of spatial cognition. Neuroscience & Biobehavioral Reviews. 2012 Apr;36(4):1179–1187. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0149763412000176.

[26] Pasqualotto A, Spiller MJ, Jansari AS, Proulx MJ. Visual experience facilitates allocentric spatial representation. Behavioural Brain Research. 2013 Jan;236:175–179. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0166432812005682.

[27] Iachini T, Ruggiero G, Ruotolo F. Does blindness affect egocentric and allocentric frames of reference in small and large scale spaces? Behavioural Brain Research. 2014 Oct;273:73–81. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0166432814004811.

[28] Jicol C, Lloyd-Esenkaya T, Proulx MJ, Lange-Smith S, Scheller M, O'Neill E, et al. Efficiency of Sensory Substitution Devices Alone and in Combination With Self-Motion for Spatial Navigation in Sighted and Visually Impaired. Frontiers in Psychology. 2020 Jul;11:1443. Available from: https://www.frontiersin.org/article/10.3389/fpsyg.2020.01443/full.