Auditory indication system for object-finding in remote collaborative assistance

Takayuki KOMODA[1], Fumina UTSUMI[2], Takayoshi YAMADA[1] and Keiichi ZEMPO[3,*]

[1]Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Tennodai, 3058573, Japan
[2]College of Engineering Systems, University of Tsukuba, 1-1-1 Tennodai, 3058573, Japan
[3]Faculty of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Tennodai, 3058573, Japan

Abstract

In this paper, we propose a remote collaboration system to assist the person with visually impaired in object-finding. The system consists of a 360-degree image centering on a person with visually impaired and presented as a panoramic image on the PC screen of a supporter in a remote location. By clicking on the PC screen, the supporter can present the AR audio (auditory indicator) superimposed on the real space that the person with visually impaired perceives, using spatial sounds. Auditory indicator enables the person with visually impaired to understand the location of the object intuitively. We conducted experiments to clarify the effect of the proposed system on the performance time of the object-finding task and the phrases of the supporters. The results of the experiment showed that the auditory indicator enabled the supporter to guide the simulated person with visually impaired by using demonstrative pronouns such as “this” and “here”.

Keywords

Augmented reality audio, Human Augmentations, Assistive technology, Remote collaboration

1. Introduction

According to 2012 World Health Organization (WHO) report, there are approximately 285 million person with visually impaired (PVI) worldwide[1]. PVI suffer many inconveniences in their daily lives due to their inability to recognize visual information. Various assistive technologies have been studied, such as navigation aids [1, 2] and object-finding aids [3, 4, 5]. Assistive technologies have made progress in assisting PVI. On the other hand, there are still many problems before assistive technology can be widely used in daily life, such as system error rates and communication speed. Remote sighted assistance (RSA) [6] has received lots of attention in addressing these issues. RSA combines assistive technologies with remote assistance from sighted people, and systems such as VizWiz [5] and Be My Eyes[2] have been developed.

However, these remote collaboration systems do not consider the differences in spatial perception between PVI and sighted people. According to Tsai et al. [7], PVI and sighted people perceive space differently. For example, when describing a route between specific points in language, sighted people describe the route based on their location. In contrast, PVI describes the route based on the locations of landmarks. Furthermore, for route descriptions between specific points, audio navigation created by PVI is subjectively more satisfactory for PVI than that created by sighted people [8]. The reason was that PVI felt more secure, oriented, and clear when the direction of travel was explained using landmarks. In addition, in conversation among sighted people, they can point to an arbitrary area using demonstrative pronouns such as "there" and communicate without redundant expressions [9]. On the other hand, the PVI are less likely to use demonstrative pronouns in conversation because they cannot recognize visual information. They tend to communicate more verbosely compared to the sighted [7]. Redundant communication has been suggested to be a stress factor in remote collaboration [10].

According to Kraut et al. [11], in a remote collaboration among sighted people, sharing the local user’s visual space with the supporter in a remote location caused the supporter to utter demonstrative pronouns. Gupta et al. [12] also showed that sharing the local user’s visual space and the remote supporter’s gaze shortens the local user’s task performance time in a remote collaboration among sighted people. The supporter used more demonstrative pronouns than when only the visual space was shared.

In this paper, we propose an auditory indication system for remote collaboration with PVI in object-finding, which utilizes AR audio superimposed on the real space that PVI perceives. We conducted experiments to clarify the proposed system’s effect on the performance time of the object-finding task and the phrases of supporters.

APMAR’22: The 14th Asia-PacificWorkshop on Mixed and Augmented Reality, Dec. 02-03, 2022, Yokohama, Japan
*Corresponding author.
zempo@iit.tsukuba.ac.jp (K. ZEMPO)
0000-0002-4012-4417 (T. YAMADA); 0000-0003-2339-5298 (K. ZEMPO)

[1]https://www.emro.who.int/control-and-preventions-of-blindness-and-deafness/announcements/global-estimates-on-visual-impairment.html
[2]https://www.bemyeyes.com/

The contributions of this paper are as follows

- Proposal of an interface using AR audio.
- AR audio technology for remote collaboration.
- Suggestion of a new interaction between PVI and sighted people using AR audio.

2. Related work

This chapter describes studies on object-finding assistance for PVI. Kaul et al. [3] developed a mobile application that combines an object detection framework and spatial sound to assist PVI in navigation and object-finding. The application recognizes objects around a PVI using a smartphone's camera function and provides auditory feedback by spatial sound to explain the scene. The user evaluation of the auditory feedback was favorable. However, some issues related to the system, such as low object detection accuracy and narrow detection range, were mentioned.

In order to improve the reliability and usefulness of assistive technology, Bigham et al. developed VizWiz [5], which combines the assistance of sighted people. PVI uses a smartphone's camera function to take pictures, ask questions to online supporters, and receive voice responses. The system does not make object detection errors, but PVI requires some recognition of the object's location to be found. Therefore, if the object's location is unknown, it is not easy to use the system.

Be My Eyes[2] is a mobile application that allows PVI to ask for assistance from a remote location using a video chat on their smartphones. PVI does not need to be aware of the object's location in advance to receive assistance. However, the field of view that can be shared with the user is limited due to smartphone camera use. Jones et al. [13] show that in remote collaboration using smartphones, users who receive video sharing from their smartphones intentionally use information from the camera images when asking questions. However, the narrow field of view and the inability to control the direction of the camera was found to be stress factors. In Be My Eyes, since the person being assisted is PVI, the supporter cannot easily convey the information obtained from the visual images, which may cause redundant communication. Wang et al. [10] suggest that redundant communication is a stress factor in remote collaboration.

3. System design

3.1. System overview

The system proposed is shown in Figure 1. The proposed system consists of PVI, who receives support, and a supporter (sighted people) who provides support from a remote location. A 360-degree camera is used to present a panoramic image of the environment around PVI on the screen of the supporter's PC. When the supporter clicks any point in the panoramic image, auditory indicators are presented to PVI from the direction corresponding to the panoramic image. By presenting auditory indicators with spatial sounds, PVI can intuitively perform object-finding.

3.2. Interface

In this section, we describe the configuration of the interface. PVI wears a 360-degree camera (RICOH THETA Z1) on the top of the head and presents a panoramic image of the surrounding environment to the screen of the supporter's PC. PVI wears open-ear headphones (Sony LinkBuds) to listen to auditory indicators without interfering with the environmental sounds. The supporter finds the object (target) from the presented images of the surroundings of PVI and clicks on the screen. The sound source is placed on the computer's three-dimensional space corresponding to the target's position in real space by clicking on the screen. Based on the positional relationship between PVI and the sound source in the computer's three-dimensional space, spatial sounds are generated and presented to PVI.

3.3. Auditory indicator

The proposed system presents auditory indicators as spatial sounds to PVI.

Spatial sound means that meta-information such as distance, direction, and spatial extent is represented in the sound reproduction. The perceived information, such as distance and direction, is called sound image localization. Sound image localization can be applied to a sound source by giving volume, time, and frequency response differences to the left and right voices. In this paper, to present spatial sound, sound image localization is convolved with a sound source using Unity and Steam Audio.

4. Experiment

In order to clarify the effect of using the proposed system on the performance time of the object-finding task and the phrases of the supporter, we conducted an object-finding task experiment with a simulated person with visually impaired (SPVI) and a participant playing the role of a supporter in a remote location, based on the experiment by Kual et al. [3]

Eight Japanese university students were randomly paired, with one participant as SPVI and the other as the supporter. SPVI wore an eye mask. Of the four groups of eight participants in the experiment, two groups of four

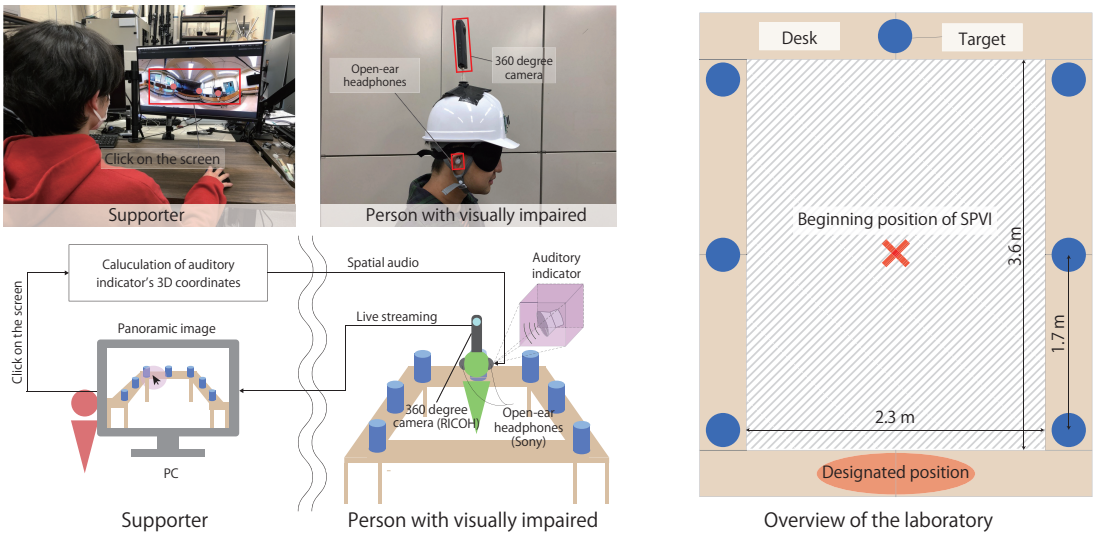

Figure 1: System configuration and overview of the laboratory.

used the proposed system, which enables the sharing of 360-degree images, auditory indicators, and voice instructions (Condition proposed.) The remaining two groups of four shared the 360-degree images and performed the task by voice instructions without auditory indicators (Condition control.).

In the laboratory, desks are arranged in four directions around the SPVI, and seven objects (targets) are placed on the desks. Figure 1 shows an overview of the laboratory.

In the experiment, SPVI is instructed by the experiment supervisor on what to find (targets). Although there are seven possible targets, SPVI is only informed of them once the experiment supervisor instructs SPVI to look for them. SPVI cooperates with the supporter through the system to find the target and carry it to the designated position. Three trials were conducted per pair. Since the eye mask blocked SPVI's vision, the experiment was conducted with SPVI sitting on a swivel chair with wheels to avoid the risk of falling.

5. Results

5.1. Performance time

The task performance time was compared between the conditions in which spatial sounds were presented and those in which they were not. The results are shown in Table 1.

Participants in Pairs #1 and #2 were presented with auditory indicators, while those in Pairs #3 and #4 were not presented with auditory indicators.

The remote collaboration could not be performed correctly in the third trial of pair #1 due to the problem with the output of the panoramic image. Therefore, the task performance time was excluded as an outlier. In addition, the task of the third trial of pair #3 was also excluded from matching the number of tasks.

As a result of the experiment, the task performance time tends to be shortened when auditory indicators are presented.

5.2. Effect on speech

As a result of the experiment, it was confirmed that when the auditory indicator was not presented, the supporters used directional phrases such as "right/left/straight" to support SPVI.

When auditory indicators were presented, phrases such as "right/left/straight" were also confirmed. On the other hand, we could confirm the use of demonstrative pronouns such as "this" and "here". In addition, it was confirmed that SPVI responded correctly to the direction of the target when demonstrative pronouns were used. Table 2 shows the number of times demonstrative pronouns, the phrases "left," "right," and "straight" were used.

5.3. Discussion

The experiment results showed that the use of auditory indicators shortened the performance time of the object-finding task and enabled the use of demonstrative pronouns by the supporters.

Table 1
Performance time. #1, #2 are using auditory indicator, #3, #4 are not using auditory indicator.

	Performance time[s]			
Pair	1st	2nd	3rd	Ave
#1	56.3	41.0	-	48.7
#2	72.6	57.5	40.7	56.9
#3	52.0	74.3	-	63.2
#4	99.6	55.3	52.8	69.2

Table 2
Number of times demonstrative pronouns, the phrases “left”, “right” and “straight” is used. Condition proposed: Using auditory indicator, condition control: Not using auditory indicator.

	Cond.: proposed			Cond.: control		
	Total	Ave	Var	Total	Ave	Var
Pro.	7	1.4	1.8	0	0	0
Left	3	0.6	0.8	7	1.4	0.8
Right	5	1.0	3.0	7	1.4	0.3
Straight	4	0.8	1.7	7	1.4	0.8

We conclude that the reason for the shorter task performance time is that the auditory indicator enables an intuitive presentation of the target direction. In the remote collaboration between sighted people, task efficiency was improved by sharing the gaze of the supporter by pointing [12]. We consider that a similar mechanism is responsible for shorting task performance time. We also note that demonstrative pronouns were used. Previous work [12, 11] confirmed that sharing the visual space enables supporters to use demonstrative pronouns, improving task efficiency. The results of this paper are consistent with these results.

The use of demonstrative pronouns is considered to be because the auditory indicators functioned in the same way as pointing for the sighted people, and a pseudo-visual space was shared between SPVI and the supporters. This is concluded from the fact that in the study of remote collaboration among sighted people by Kraut et al. [11], the supporter started to use demonstrative pronouns after the visual space was shared.

6. Conclusion and future works

In this paper, we propose a system that uses auditory indicators to provide directions to assist PVI in object-finding through remote collaboration. In order to clarify the effectiveness of the proposed system, we conducted an object-finding task experiment. We investigated the effect of the auditory indicator on the task performance time and the phrases of supporters. As a result of the experiment, it was confirmed that the task performance time tended to be shortened when auditory indicators were presented. In addition, it was also confirmed that the auditory indicator made the supporter use more demonstrative pronouns such as "this" and "here". We concluded that these results are because the presentation of the target direction by auditory indicators functioned in the same way as gaze sharing in the remote collaboration between sighted people. The results of this paper suggest that demonstrative pronouns can be used in remote collaboration with PVI, indicating a new interaction between PVI and sighted people using AR audio. However, one limitation of this paper is that the participants of the experiment were SPVI. Since SPVI is a blindfolded sighted person, it is not representative of PVI, and future studies should be conducted with PVI as a participant.

In the future, it is necessary to investigate whether the presentation of auditory indicators has the same role as gaze sharing and to clarify the mechanism by which the demonstrative pronouns were used.

References

[1] M. H. A. Wahab, A. A. Talib, H. A. Kadir, A. Johari, A. Noraziah, R. M. Sidek, A. A. Mutalib, Smart cane: Assistive cane for visually-impaired people, arXiv preprint arXiv:1110.5156 (2011).

[2] A. Helal, S. E. Moore, B. Ramachandran, Drishti: An integrated navigation system for visually impaired and disabled, in: Proceedings fifth international symposium on wearable computers, IEEE, 2001, pp. 149–156.

[3] O. B. Kaul, K. Behrens, M. Rohs, Mobile recognition and tracking of objects in the environment through augmented reality and 3d audio cues for people with visual impairments, in: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, 2021, pp. 1–7.

[4] M. Eckert, M. Blex, C. M. Friedrich, et al., Object detection featuring 3d audio localization for microsoft hololens, in: Proc. 11th Int. Joint Conf. on Biomedical Engineering Systems and Technologies, volume 5, 2018, pp. 555–561.

[5] J. P. Bigham, C. Jayant, H. Ji, G. Little, A. Miller, R. C. Miller, R. Miller, A. Tatarowicz, B. White, S. White, et al., Vizwiz: nearly real-time answers to visual questions, in: Proceedings of the 23nd annual ACM symposium on User interface software and technology, 2010, pp. 333–342.

[6] S. Lee, M. Reddie, C.-H. Tsai, J. Beck, M. B. Rosson, J. M. Carroll, The emerging professional practice of remote sighted assistance for people with visual impairments, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1–12.

[7] K.-Y. Tsai, Y.-H. Hung, R. Chen, E. Chang, Indoor

spatial voice navigation for people with visual impairment and without visual impairment, in: Proceedings of the 2019 7th International Conference on Information and Education Technology, 2019, pp. 295–300.

[8] Y.-H. Hung, K.-Y. Tsai, E. Chang, R. Chen, Voice navigation created by vip improves spatial performance in people with impaired vision, International journal of environmental research and public health 19 (2022) 4138.

[9] Y. Hato, S. Satake, T. Kanda, M. Imai, N. Hagita, Pointing to space: modeling of deictic interaction referring to regions, in: 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, 2010, pp. 301–308.

[10] B. Wang, Y. Liu, J. Qian, S. K. Parker, Achieving effective remote working during the covid-19 pandemic: A work design perspective, Applied psychology 70 (2021) 16–59.

[11] R. E. Kraut, D. Gergle, S. R. Fussell, The use of visual information in shared visual spaces: Informing the development of virtual co-presence, in: Proceedings of the 2002 ACM conference on Computer supported cooperative work, 2002, pp. 31–40.

[12] K. Gupta, G. A. Lee, M. Billinghurst, Do you see what i see? the effect of gaze tracking on task space remote collaboration, IEEE transactions on visualization and computer graphics 22 (2016) 2413–2422.

[13] B. Jones, A. Witcraft, S. Bateman, C. Neustaedter, A. Tang, Mechanics of camera work in mobile video collaboration, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015, pp. 957–966.