Communication Support System for Deaf and Hard of Hearing People by Captions Considering Sound Source and Sound Direction

Nobuko Kato[a] and Marie Kepler
Faculty of Industrial Technology, The National University Corporation of Tsukuba University of Technology, Tsukuba City, Japan

Keywords: Communication Support, Sound Environment, Sound Awareness, Deaf and Hard of Hearing.

Abstract: For a deaf and hard of hearing (DHH) visitor to be able to understand all information in a museum, it is necessary to present not only the human voice as captions, but also the sounds in the environment in such a way that the DHH visitor can understand. For this reason, we investigated communication support methods in a way that not only presents speech content as captions but also presents sound information, such as the source and direction of the sound source. In this paper, we examine a case in which the direction of the sound source is indicated by flowing text, and another case where the sound source is indicated not only by text but also by photos and pictograms, using the following two presentation methods: wearable AR glasses and floor projection.

1 INTRODUCTION

Although museums are important places for learning, there are many situations in which there is a lack of information for the deaf and hard of hearing (DHH) (Namatame et al., 2020). These include problems such as the inability to grasp the information that a hearing person would be able to obtain. Therefore, speech recognition and text interpretation systems are widely used to improve communication accessibility and solve such problems (Wakatsuki et al., 2018). Usually, these systems convert only the speech of the nearest speaker into text and display it as a caption without displaying information about other sounds.

However, in real life, there are various sounds other than human speech, such as the sound of an ambulance siren or knock at the door, which are important sources of information. This is also true for museums, aquariums, and zoos. It has been pointed out that sound itself is an important element of the exhibition experience (Orhan, 2019; Shuko, 2003). In museums, there are sounds from speakers and other items that produce sounds, including sounds from other devices (e.g., exhibits), sounds of human activity (talking, footsteps, and work), and sounds of nature (e.g., living things and the wind), which create the unique atmosphere of a museum, such as bustle and silence. Songs and noises made by creatures and objects that can be considered as part of an exhibit and announcements in a museum are essential sounds used to pass information to visitors.

Whereas general visitors can enjoy their museum experience by obtaining information in such a soundscape, DHH visitors have difficulty obtaining information from the surrounding sounds. Even if the results of speech recognition are displayed in a text form, only a small portion of the various sound information, human voices, and sounds can be collected by the microphone. Listening to environmental sounds helps understand the source of the sound and the surrounding situation (Gaver, 1993), and it is desirable to have a communication support system that can present sound information, including the source and direction of the sound source.

In this study, to provide information accessibility suitable for DHH visitors in museums, we examine how to present sound information that includes sound information, such as the source and direction of the sound source. For this purpose, we will analyze a case in which the method of presenting text was modified, and another case in which not only text but also photographs or pictograms representing sound sources were added.

[a] https://orcid.org/0000-0003-0657-6740

Kato, N. and Kepler, M.
Communication Support System for Deaf and Hard of Hearing People by Captions Considering Sound Source and Sound Direction.
DOI: 10.5220/0011077800003182
In *Proceedings of the 14th International Conference on Computer Supported Education (CSEDU 2022) - Volume 2*, pages 675-680
ISBN: 978-989-758-562-3; ISSN: 2184-5026
Copyright © 2022 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

2 RELATED STUDIES

Studies on DHH individuals listening to environmental sounds suggest that they use a strategy of visually searching for sound sources in the background information, and it is believed that it is necessary to capture background information using visual information (Tabaru et al., 2011). Studies on listening comprehension with hearing aids and cochlear implants have reported that even cochlear implant users have difficulty distinguishing fine categories of sounds (Inverso and Limb, 2010). A system for learning environmental sounds on a PC was developed (Kato et al., 2018; Shafiro et al., 2015), and it has been found that learning environmental sounds is effective in recognizing sound sources.

A system that automatically recognizes various surrounding sounds and notifies the hearing impaired is also being considered (Matthews et al., 2006; Goodman et al., 2020; Finlater et al., 2019). Experiments on peripheral visual displays used to all deaf people to maintain awareness of nonspeech sounds in the environment showed that participants confirmed the importance of sound information notifications and sought information on location, volume, and sound identity. When a smartwatch was used to provide sound feedback to a deaf person using a combination of visual and tactile feedback, positive responses to sound identity, sound direction, and volume were obtained. Particularly when the sound in the environment was complex, all DHH persons who participated in the experiment wanted their feedback to be filtered. In other words, it is necessary to know what the sound is, including its direction, for the required interpretation.

To identify and visualize a sound source, a method for presenting the sound source on an HMD was proposed (Guo et al., 2020). Although such a method is extremely easy to understand when the sound source is within the field of view, it has problems in displaying the sound source when it is outside the field of view.

At present, captions are most widely used as a method of communication support. Therefore, in situations in which captions are used, the sound source and its direction are also expected to be displayed in combination with captions. In this paper, we examine how to present not only human voices as captions, but also present sounds in the environment in a way that DHH visitors can understand.

Table 1: Conditions of the experimental video (presentation method and information presented).

	Stationary	Flow and then stop	Flow
Text	B	E	H
Text and photo	C	F	I
Text and pictogram	D	G	J

* Condition A is the caption presented at the bottom of the screen in the conventional method.

3 EXPERIMENT 1: COMPARISON OF PRESENTATION CONTENTS

3.1 Method of Experiment 1

To determine how to present captions such that the content and direction of the sound source are understood, experiments were conducted to determine the effectiveness of the method used to present the sound from the direction of the sound source as well as the effectiveness of using non-text to present the information.

We prepared the following three of presentation methods.

Stationary: Captions are displayed in a stationary state.

Flow and then Stop: The captions move from the direction of the sound source and then stop.

Flow: Captions move from the direction of the sound source and pass by the viewer.

In addition, the following three types of information were prepared for presentation.

Text: Text only is presented.

Text and Photo: Text and photographs are presented.

Text and Pictograms: Text and pictograms are presented.

We prepared a total of 10 types of videos, including 9 types (Table 1) created by combining the 3 methods of presentation and the 3 types of information presented above. We then created a video by presenting static captions at the bottom of the screen, which is the conventional caption presentation method. The captions we created represent four types of sounds (the buzz of cicadas, frog calls, songs of Japanese bush warblers, and human speech) coming from various directions. Figure 1 shows examples of pictograms used in the experiment, and Figure 2 shows an image of a display with flowing text and a pictogram. In Japanese, the sounds of cicadas, frogs, and other

creatures are represented by onomatopoeic words.

These videos were presented to 12 DHH participants in their 20s, and they answered a corresponding questionnaire. The videos were presented on the monitor, and the order of viewing the captioned videos was determined using random numbers to avoid bias in the presentation method and the information presented.

Figure 1: Examples of pictograms used in the experiment.

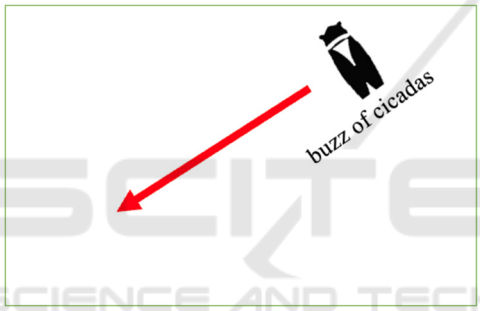

Figure 2: Image of display with flowing text and a pictogram. Red arrows indicate the direction of flow.

3.2 Experiment Results

In Experiment 1, the mean and standard deviation of the rating values (6 levels, 1 = strongly disagree, 6 = strongly agree) for the question "Was it easy to read?" are shown in Figure 3.

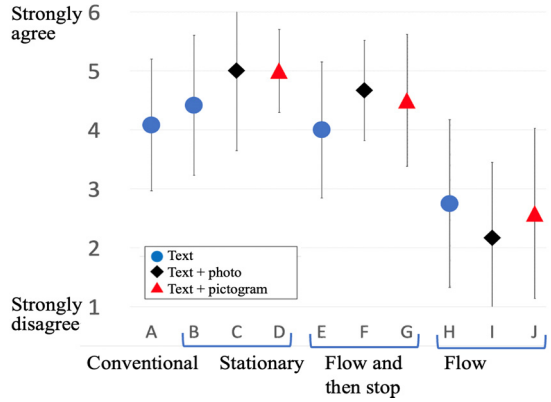

Figure 3: Questionnaire results for "Was it easy to read?".

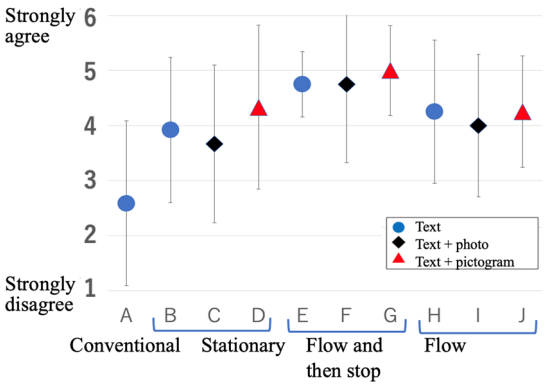

Figure 4: Questionnaire results for "Was the direction of the sound source easy to understand?".

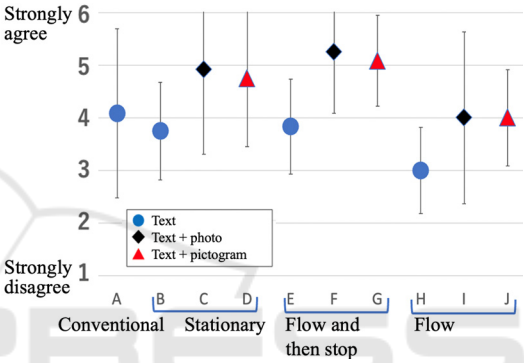

Figure 5: Questionnaire results for "Was it easy to recognize the sound identity?".

Table 2: Types of videos prepared.

	Stationary	Flow and then stop	Flow
Text	A	B	C
Text and pictogram	D	E	F

* Condition G is the caption presented at the bottom of the screen for the conventional method.

The ANOVA analysis showed a significant difference in Factor 1 (methods of presentation); therefore, the Bonferroni method was used as a sub-test. The readablity of Flow and then stop method did not differ from that of conventional captions. By contrast, the Flow method was rated significantly lower ($p < 0.05$).

Figure 4 shows the mean and standard deviation of the evaluation values (6 levels) for the question "Was the direction of the sound source easy to understand?" The presentation method in which captions flow from off-screen and stop at a certain position (Flow and then stop) or a method of presentation in which captions flow from off-screen to off-screen and disappear (Flow) was rated

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

significantly higher than conventional captions ($p < 0.05$). In other words, the Flow and then stop or Flow method is considered easier for understanding the direction of the sound source compared to conventional captions.

Figure 5 shows the results of the questionnaire for "Was it easy to recognize the sound identity?" In the case of the Flow and then stop captions, the sound identity was easier to recognize with the text + photo (F) or the text + pictogram (G) than with the text only (E) ($p < 0.05$).

4 EXPERIMENT 2: COMPARISON OF PRESENTATION DEVICES

4.1 Method of Experiment 2

To study the presentation devices used to convey the sound source clearly, we conducted an experiment to compare AR glasses with a projector that projects on the floor. Because no difference was found between photos and pictograms in Experiment 1, only text and pictograms were used to present information in this experiment.

From the four types of sounds from cicadas, frogs, Japanese bush warblers, and announcements, seven types of videos were prepared, including six created by combining three presentation methods and two types of information presentation (Table 2), one of which was created using the conventional caption posting method.

These videos were presented to DHH participants using a floor projector and AR glasses (Epson MOVERIO BT-30E), and they were asked to answer a questionnaire each time. The projection size of the projector was 150 in, and the projection size of the AR glasses was 40 in at a distance of 2.5 m, as shown in Figure 6.

4.2 Experiment Results

Figures 7 and 8 show the mean and standard deviation of the evaluation values (six levels) for projector and AR glasses, respectively. As a result of ANOVA, a difference was found in Factor 1 (presentation method); therefore, the Bonferroni method was used as a subtest. For both the projector and AR glasses, the presentation method in which the captions Flow and then stop or simply Flow had significantly higher ratings than the traditional captions ($p < 0.05$).

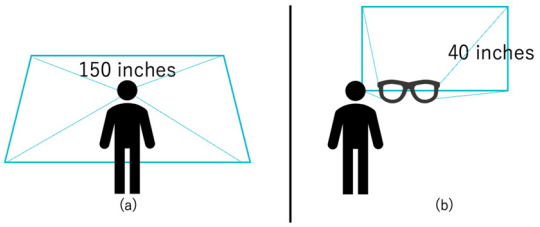

Figure 6: Layout in Experiment 2. (a) Layout for projection on the floor with a projector, (b) the projection size of the AR glasses was 40 in at a distance of 2.5 m.

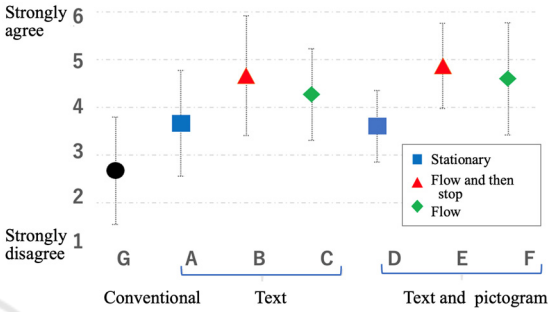

Figure 7: Results of the questionnaire on "Was the direction of the sound source easy to understand?" when using a projector.

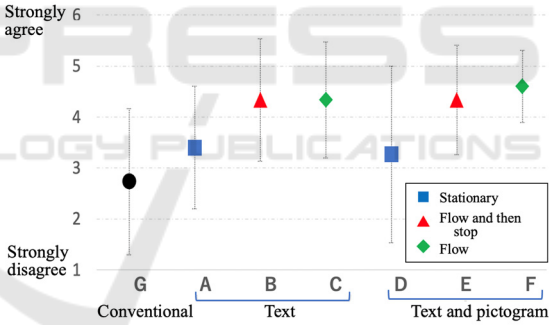

Figure 8: Results of the questionnaire on "Was the direction of the sound source easy to understand?" when using AR glasses.

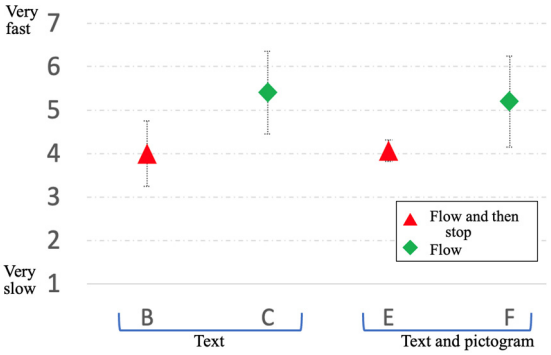

Figure 9: Results of the questionnaire on "Was the speed of the text flow appropriate?".

The mean and standard deviation of the evaluation values (7 levels, 7 = very fast, 4 = appropriate, 1 = very slow) for the question "Was the speed of the text flowing through the projector appropriate?" are shown in Figure 9. Compared with the presentation method Flow and then stop, the presentation method Flow was found to be significantly faster ($p < 0.05$)

5 DISCUSSION

From the results of Experiment 1, we found that the two methods Flow and then stop and Flow, in which the captions moved from the direction of the sound source, were better than the conventional captions in terms of clarity of the sound source direction. From the results of the questionnaire on readability, it was found that the method in which the captions moved from the direction of the sound source and passed directly outside the screen (Flow) is more difficult to read than conventional captions. In other words, a method in which the captions move from the direction of the sound source and stop (Flow and then stop) is considered effective and easy to understand and read. However, a more detailed analysis is needed, such as differences in the sound content and number of captions.

In the results of the questionnaire on sound sources, photographs and pictograms were rated higher than captions alone. When sounds are displayed as text, it is expected that DHH users, who have little experience listening to such sounds, will have difficulty recognizing what they represent. Therefore, photographs and pictograms that can be grasped intuitively are highly valued.

In Experiment 2, the direction of the sound source was easy to find, even when using a floor projector or AR glasses.

However, in the presentation method in which the captions moved and flowed as they were on the projector, it was found that the speed of the captions needed to be adjusted.

Here, we compare the cases in which AR glasses and a projector are used in a museum. With AR glasses, the relative direction of the sound source changes depending on the orientation of the individual's face. It is therefore necessary to detect the direction of the sound in real time with a small device. It is also necessary to consider how the information is displayed. However, when the floor projector is installed in the museum, the direction of the sound can be easily identified, and the relative direction of the sound source can be fixed as long as the source does not move. In addition, in a complex sound environment with various sounds, if a lot of information is to be displayed simultaneously, the AR glasses may block the field of view. In contrast, with a floor projector, the field of view of the user is maintained, and the sound information can be checked when needed. However, because it is expected to be difficult to discriminate text information in peripheral vision, actual experiments in a complex sound environment are needed.

6 CONCLUSION

To provide appropriate information for DHH visitors in a museum, we studied how to provide information that includes sound information, such as the sound source and its direction. As a result of the experiment, we found that the presentation method in which the captions flow from the direction of the sound source and stop at a certain position is effective in determining the direction of the sound source, regardless of the device. In addition to text, information such as pictograms was also effective.

Experiments presented on AR glasses and a floor projector also showed that the proposed method for displaying captions contributed to the clarity of the direction of the sound source. In the case of a floor projection with a projector, the captions flowing from the direction of the sound source to the outside of the screen were said to be displayed "too fast," and the speed had to be examined. Previous studies have shown that large displays are preferred indoors for non-speech notifications (Matthews et al, 2016), and the floor projection method using a projector is considered to be one of the most effective methods for application in a museum because of its ease in identifying the sound direction and convenience of use.

It is expected that the amount of sound information displayed and the length of the captions will also affect the evaluation. Therefore, we would like to further study ways to present captions with high readability such that the sound source and its direction can be identified.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Numbers JP18H03660.

REFERENCES

Namatame, M., Kitamura M, Iwasaki, S. (2020). The science communication tour with a sign language interpreter. *Pacific Rim International Conference on Disability and Diversity Conference Proceedings*, Center on Disability Studies, University of Hawai'i at Mānoa.

Wakatsuki, D., Kato, N., Shionome, T., Kawano, S., Nishioka, T., and Naito, I. (2017). Development of web-based remote speech-to-text interpretation system captiOnline. *Journal of Advanced Computational Intelligence and Intelligent Informatics,* 21(2), 310–320. 10.20965/jaciii.2017.p0310

Orhan, C. (2019). *A comparative study on indoor soundscape in museum environments*, Thesis (M.S.), Bilkent University, http://hdl.handle.net/11693/52316

Shuko, K. (2003). Analysis of the exhibition method in "exhibition on the theme of sound" of museums, *Cultural Information Resources*, 10(2).

Gaver, W. W. (1993). What in the world do we hear?: An Ecological approach to auditory event perception. *Ecological Psychology*, 5(1), 1–29. 10.1207/s15326969eco0501_1.

Tabaru, K., Harashima, T., Kobayashi, Y., and Katada, A. (2011). Effects of aided audible frequencies and contextual information on identification of environmental sounds by individuals with hearing impairments: Analysis of individual cases. *The Japanese Journal of Special Education*. 48(6), 521–538. 10.6033/tokkyou.48.521.

Inverso, Y. and Limb, C. J. (2010). Cochlear implant-mediated perception of nonlinguistic sounds. *Ear and Hearing*. 31(4), 505–514. 10.1097/AUD.0b013e3181d99a52.

Kato, Y., Hiraga, R., Wakatsuki, D., and Yasu, K. (2018). A preliminary observation on the effect of visual information in learning environmental sounds for deaf and hard of hearing people. *ICCHP 2018, Proceedings* (1), 183–186. 10.1007/978-3-319-94277-3_30.

Shafiro, V., Sheft, S., Kuvadia, S., and Gygi. B. (2015). Environmental sound training in cochlear implant users. *Journal of Speech, Language, and Hearing Research*. 58(2), 509–519. 10.1044/2015_JSLHR-H-14-0312

Matthews, T., Fong, J., Ho-Ching, F.W.L., and Mankoff, J. (2006). Evaluating non-speech sound visualizations for the deaf. *Behaviour & Information Technology*. 25(4), 333–351. 10.1080/01449290600636488.

Goodman, S., Kirchner, S., Guttman, R., Jain, D., Froehlich, J., and Findlater, L. (2020). Evaluating smartwatch-based sound feedback for deaf and hard-of-hearing users across contexts. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Computing Machinery*, 1–13. 10.1145/3313831.3376406.

Findlater, L., Chinh, B., Jain, D., Froehlich, J., Kushalnagar, R., and Lin. A. C. (2019). Deaf and hard-of-hearing individuals' preferences for wearable and mobile sound awareness technologies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Paper 46, 1–13. 10.1145/3290605.3300276.

Guo, R., Yang, Y., Kuang, J., Bin, X., Jain, D., Goodman., Findlater, L., and Froehlich, J. (2020) HoloSound: Combining speech and sound identification for deaf or hard of hearing users on a head-mounted display. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20)*, Article 71, pp. 1–4. 10.1145/3373625.3418031.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS