



See-Through Captions in a Museum Guided Tour: Exploring Museum Guided Tour for Deaf and Hard-of-Hearing People with Real-Time Captioning on Transparent Display

Ippei Suzuki^(✉), Kenta Yamamoto, Akihisa Shitara, Ryosuke Hyakuta, Ryo Iijima, and Yoichi Ochiai

Research and Development Center for Digital Nature, University of Tsukuba, Tsukuba, Japan

contact@digitalnature.slis.tsukuba.ac.jp, wizard@slis.tsukuba.ac.jp

Abstract. Access to audible information for deaf and hard-of-hearing (DHH) people is an essential component as we move towards a diverse society. Real-time captioning is a technology with great potential to help the lives of DHH people, and various applications utilizing mobile devices have been developed. These technologies can improve the daily lives of DHH people and can considerably change the value of audio content provided in public facilities such as museums. We developed a real-time captioning system called See-Through Captions that displays subtitles on a transparent display and conducted a demonstration experiment to apply this system to a guided tour in a museum. Eleven DHH people participated in this demonstration experiment, and through questionnaires and interviews, we explored the possibility of utilizing the transparent subtitle system in a guided tour at the museum.

Keywords: Real-time captioning · Deaf and hard-of-hearing · Transparent display · Accessibility · Museum · Guided tour

1 Introduction

Deaf and hard-of-hearing (DHH) people have difficulty obtaining auditory information. There are several technologies that have been developed to address this problem, including hearing aids, cochlear implants, and subtitles. Furthermore, as a social issue, various efforts and technologies are becoming more common to ensure accessibility for persons with disabilities.

Automatic speech recognition (ASR) is a typical example of a technology for people with deafness. ASR has long been considered a universal access method for audio information [21]. The introduction of such technology has

I. Suzuki and K. Yamamoto—These authors contributed equally to this research.

© The Author(s) 2022

K. Miesenberger et al. (Eds.): ICCHP-AAATE 2022, LNCS 13341, pp. 542–552, 2022.

https://doi.org/10.1007/978-3-031-08648-9_64



Fig. 1. Overview of the See-Through Captions device for use in museum guided tours. The two people on the left are visitors, and the person on the right is a guide.

been attempted in the field of education [1], and user studies have been conducted to explore how DHH people use and benefit from ASR [6, 7, 9]. Research has also been conducted on DHH people to freely utilize speech recognition in more varied settings than the classroom [12, 22]. In recent years, research into text conversion using ASR on mobile devices has increased [5, 10], and studies have been actively conducted on the use of augmented reality (AR) devices for displaying captions [4, 16, 17].

However, there are restrictions specific to the method of real-time captioning by ASR using such mobile and AR devices. For example, when communicating using a mobile device, the facial expressions of the conversation partner cannot be properly attended to by the DHH person because the mobile device must be viewed to see the result of speech recognition. When using an AR device, DHH people can see the speech-recognized text while looking at the other person's facial expressions. However, the speaker cannot confirm whether the system has misrecognized their words, and it may occur a discrepancy in communication.

The importance of accessibility to such speech information is being examined in areas other than educational settings, such as the museum. There has been considerable research describing the importance of improving information accessibility for sensory-impaired people in museums [8, 11]. For DHH people, audible information is most difficult to access in museums. Hence, sign language-guided tours are often offered [14]. Alternatively, mobile devices have also been utilized for displaying auditory information to DHH people [2, 13, 18]

However, these methods have various problems. First, a guided tour with a sign language interpreter has the problem that it is difficult to recruit an interpreter. Sign language translators improve the quality of information that DHH people can receive, but at a higher social and financial cost. Second, the information presented on mobile devices is a one-way information transmission method. Although users can read the information easily when using such a system, they cannot communicate with the guide on the museum tour.

To deal with these problems, we have developed a handheld type of See-Through Captions, a real-time captioning system that utilizes a portable transparent display and allows conversational partners to confirm captions without interfering with nonverbal communication, as shown in Fig. 1. We discussed findings based on the results of a case study in which DHH people actually participated in a guided tour of a museum using See-Through Captions.

2 See-Through Captions for Guided Tours

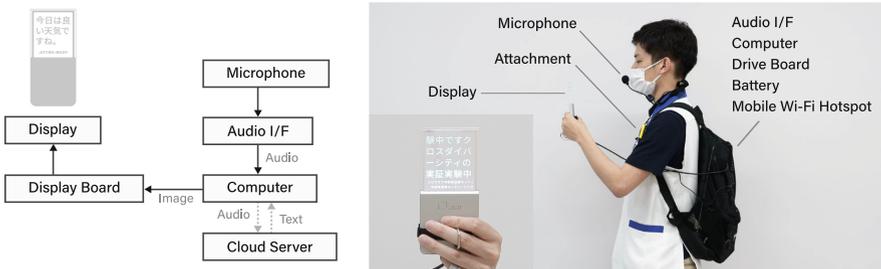


Fig. 2. System configuration of See-Through Captions.

See-Through Captions is a system that displays the results of real-time captions via ASR on a transparent display [23]. In this study, the system was downsized and made portable such that it could be used during a guided tour in a museum, as shown in Fig 2. First, we used a small, portable transparent display with a length of 8 cm and a width of 7 cm. As this prototype, Japan Display Inc.’s transparent display was used [15]. Projected images on this transparent display can be seen from both sides. The resolution of the display was 320×360 pixels. The weight of the display was approximately 130 g. Next, a headset microphone (WH20XL; Shure Inc.) was used for speech input. The computer that performed the speech recognition processing, the drive board of the display, the audio interface, the mobile Wi-Fi hotspot, and the battery were included in the backpack. The total weight of the backpack was approximately 3.3 kg. For the speech recognition process, the user inputs speech to the headset microphone, and the speech data are processed on the cloud server via the Web Speech API¹ on a web browser (Google Chrome; Google LLC).

¹ https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API [Last access date: Apr. 1, 2022].

3 Case Study: Guided Tour in Museum

As a case study in a museum, we conducted a guided tour using See-Through Captions. We collaborated with Miraikan - The National Museum of Emerging Science and Innovation, Japan. Miraikan hosts tours for their exhibitions by science communicators. Guided tour programs using See-Through Captions were planned by discussing between authors and science communicators. Guided tours were conducted in the Japanese language.

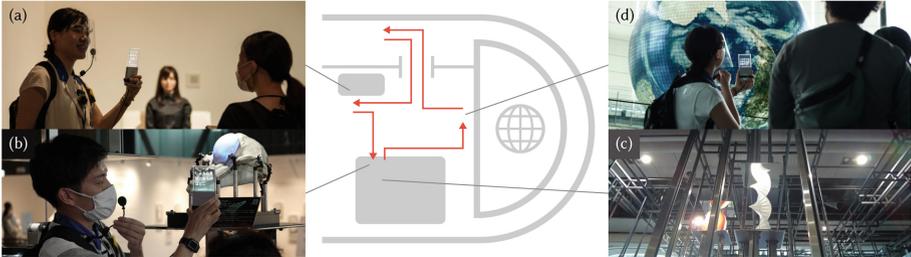


Fig. 3. Contents of the guided tour. The map depicts the third floor of Miraikan. The red arrow shows the route of the guided tour. (Color figure online)

3.1 Study Design

Tour Contents. Tour contents were designed under the following preconditions: only one group could participate in one tour, each group was required to contain at least one DHH person, a guide person used See-Through Captions when they spoke, and communication from DHH people to the guide person was through speech or writing. And the guide described the communication protocol of the tour: the guide would express “wait” in gestures of sign language if the ASR system stopped while the guide was talking, participants would raise their hand or notepad when they wanted to talk, and participants would express “applause” in gestures of sign language when someone talked one’s idea. The guide described the theme of the tour and conducted some quiz games about Miraikan. After this brief introduction, they entered the exhibition area.

The tour guide explained the four exhibits. Figure 3 shows the route of the tour and the appearance of each exhibit. The theme of this tour was “The difference between humans and robots”. Figure 3(a) shows an exhibition of moving androids (human-like robots). Figure 3(b) shows the exhibition of a dolphin’s echolocation mechanism using sound and light. Figure 3(c) shows the exhibition of a DNA Origami’s structural model. Figure 3(d) depicts the Geo-Cosmos, which is a “globe-like display” showing images of clouds. The exhibition depicted in Fig. 3(c) was excluded for some tour groups, depending on the tour’s progress and other scheduled events.



Fig. 4. Methods of See-Through Captions use in guided tour: (a) basic position, (b) overlay position, (c) hands-free position.

How to Use See-Through Captions. There are three ways to use See-Through Captions during a guided tour, as shown in Fig. 4. Figure 4(a) depicts the most basic way to use See-Through Captions. The display was placed in front of the guide’s face. Figure 4(b) depicts the method in which the transparent display was overlaid in front of the exhibit so that participants could see the linguistic information while looking at the exhibit. This enabled the guide to use demonstrative pronouns such as “this” or “here” while pointing at a specific position of exhibits. Figure 4(c) depicts the method wherein the display was fixed to the chest attachment so that the guide was able to communicate hands-free. This position enabled the guide to make hand movements. For example, in Fig. 4(c), the guide describes how we can express the “International Space Station” in gestures of the Japanese sign language. Through the guided tour, the guide alternated between these three usages flexibly.

Questionnaire and Interview. We created questionnaires about the usability of See-Through Captions. The questionnaires included questions on the following: (1) the readability of the ASR results, (2) the noticeability of misrecognition of ASR, and (3) whether they wanted to continue utilizing such a device. They were rated on a 5-point Likert scale. In addition, we asked the following topics with free description questions: (4) “if you would like to continue utilizing the device, which situation in your daily life would you like to use it?”, and (5) “are there any inconvenience points or improvements you think could be made to the device?”.

3.2 Procedure

The study was conducted in a permanent exhibition of Miraikan. Each participant was briefly informed of the purpose of the study and told that they could exit the experiment at any time. These explanations were provided by pre-recorded videos with sign language and open captions. Participants were provided with a consent form to sign. They were then asked about the preferred position of See-Through Captions and preferred infection-prevention methods (face shield or face mask). After the completion of the guided tour, the participants were asked to fill out the questionnaires and be interviewed about the

guided tour and their answers to the questionnaires. The total time required for the entire process, including one tour and interview, was approximately 60 to 90 min. This study was approved by the research ethics committee of the Faculty of Library, Information and Media Science, University of Tsukuba.

3.3 Participants

Each tour group contained at least one DHH person; some groups contained a few hearing people. We conducted nine guided tours in this study. Seven of them included one DHH person, and the others included two DHH persons. Three groups included one hearing person, and one group included two hearing persons. There are eleven DHH participants (7 females and 4 males), aged between 18 and 53 years ($M = 38$, $SD = 10.9$), four hearing participants (3 females and 1 male), aged between 36 and 56 years ($M = 45.8$, $SD = 7.8$), and one hearing participant without questionnaires. Nine DHH participants had a profound impairment, including deafness, one DHH participant had a severe impairment, and another DHH participant had not answered about one's impairment. This classification is based on the WHO's criteria for hearing impairment [3]. We recruited participants by posting on the Miraikan website and some social network services.

3.4 Qualitative Evaluation

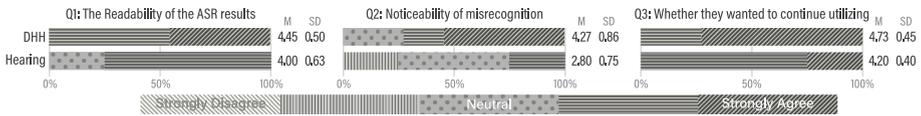


Fig. 5. Result of questionnaires (5-point Likert scale).

The results of questions that can be answered quantitatively are presented in this section. The aggregated results of the scores of DHH people and hearing people are illustrated in Fig. 5. First, the readability of the ASR results was highly evaluated by all DHH people and by all but one hearing person. Next, regarding whether it was easy to notice misrecognition in the text, some people responded that it was very easy, whereas other respondents did not find it easy. In particular, one hearing person found it difficult to notice instances of misrecognition. Finally, all participants responded positively to the question of whether they would like to continue using the device in the future. Taken together, it is interesting to note that DHH people gave higher marks to this system than hearing people did.

3.5 Quantitative Evaluation

We asked the participants to freely describe their answers to questions (4) and (5), and conducted further interviews. This section summarizes the pros/cons and issues of our method.



Fig. 6. Relationship between ruby, kanji, and kana in Japanese.

Automatic Speech Recognition. Because the core of this system is ASR, knowing how to interact with ASR is important. Among the functions we prepared, although ruby (giving kana above kanji to show the letters' readings, as shown in Fig. 6) was well received, many people identified that it was difficult to read when misrecognition occurred. However, such an accident is a fundamental issue of ASR, and it is difficult to completely overcome it. In the interviews, with this background in mind, participants often responded that it would be better for the speaker to acquire utterances and speaking styles that were easy for the system to recognize correctly. For example, speaking without using a double negation or writing technical terms, and proper nouns in a separate panel. In addition, many people cited dictionary registration as a function to be expected in the future. This is because it was often difficult to identify technical terms and proper nouns, and that such technical words are to be predicted in contexts such as museum-guided tours.

Readability of Captions. The readability is easily affected by the background color/scenery and the reflection on the display. Although the system provided functions of changing the text and background colors, the white color that was felt to be the most visible in the authors' pretest was applied, and there was no background color. As a result of conducting a guided tour, many participants commented that it might be difficult to see in some settings (especially when there is a strong light in the background). In addition, there was an opinion that it is important to be careful because the reflection on the display changes depending on the weather. Because these points cannot be dealt with entirely through system design, it will be important for the guide to be careful with their behavior and have a system that allows easy changes to the design.

How to Display Captions. In this experiment, the subtitle display method was basically unchanged from the initial stationary type [23], and was originally suitable for large display sizes. Therefore, in the small version, there was an

opinion that the character flow was too fast and that the screen was filled with rephrasing when misrecognition occurred. There was also an opinion that the small size of the screen limited the number of subtitles that could be displayed at the same time; hence, the system should have a function to look back at the history of the conversation.

Benefits of Transparency. We were able to obtain many positive opinions from the participants that they could see the subtitles while looking at the contents of the exhibition. There was also an opinion that it was easy to communicate in both directions by being able to see the guide's face and make eye contact. Other studies have also confirmed that DHH people tend to prefer eye contact when communicating [19,20]. In addition, some participants said that transparency made it possible to see the whole without obstructing the view, and that they did not feel any gap.

Display Position. Because this was a handheld setup, it was relatively easy to change the position of the device. In addition, when starting the guided tour, we confirmed a position that was easy for the participants to see. Based on this premise, in the interview, there was an opinion that "if the display is held near the face, it is easier because there is only one place to watch."

Display Type and Size. The handheld See-Through Captions used in the guided tour have features that are not found in existing displays. Therefore, it is necessary to compare it with other methods in detail in the future. In comparison with other methods, the opinion obtained in this experiment was that although the use of AR glasses was tiring, the See-Through Captions system was easier. Alternatively, there were many opinions that the display itself was small, and there were many complaints about the number of line breaks due to the small size of the screen.

Challenges Specific to Guided Tours. See-Through Captions was originally developed as a one-to-one communication support system [23]. This experiment was the first to introduce the system to a guided tour. While there were many positive opinions, issues specific to guided tours were also found. For example, if this system is used with the guide's mouth visible, people will try to read lips and understand the presented linguistic information at the same time, which can cause confusion. When multiple people participate in the tour, it will be easier to have more fulfilling communication if the audio content of the participants is also displayed.

4 Conclusions

In this study, we conducted experiments using See-Through Captions for guided tours in a museum. While many of the evaluations of our system were well received, some issues have remained. In particular, we must carefully consider the means of information transmission from DHH people. The current system assumes that DHH people speak using voice; however, some DHH people do not tend to speak by their voice. To better accommodate communication with these people using our system, it is necessary to consider an additional input interface.

Acknowledgements. This work was supported by JST CREST Grant Number JPMJCR19F2, including the AIP Challenge Program, Japan. We would like to thank Japan Display Inc. for lending us the prototype of the transparent display, and also to thank Miraikan - The National Museum of Emerging Science and Innovation (especially Bunsuke Kawasaki, Sakiko Tanaka, and Chisa Mitsuhashi) for their unfailing support and assistance.

References

1. Bain, K., Basson, S.H., Wald, M.: Speech recognition in university classrooms: Liberated learning project. In: Proceedings of the Fifth International ACM Conference on Assistive Technologies, Assets 2002, pp. 192–196. Association for Computing Machinery, New York (2002). <https://doi.org/10.1145/638249.638284>
2. Constantinou, V., Loizides, F., Ioannou, A.: A personal tour of cultural heritage for deaf museum visitors. In: Ioannides, M., et al. (eds.) EuroMed 2016. LNCS, vol. 10059, pp. 214–221. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48974-2_24
3. On Prevention of Deafness, I.W.G., Hearing Impairment Programme Planning: Geneva, S., for the Prevention of Deafness, W.H.O.P., Impairment, H.: Report of the informal working group on prevention of deafness and hearing impairment programme planning, Geneva, 18–21 June 1991 (1991)
4. Jain, D., Chinh, B., Findlater, L., Kushalnagar, R., Froehlich, J.: Exploring augmented reality approaches to real-time captioning: a preliminary autoethnographic study. In: Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems, DIS 2018 Companion, pp. 7–11. Association for Computing Machinery, New York (2018). <https://doi.org/10.1145/3197391.3205404>
5. Jain, D., Franz, R., Findlater, L., Cannon, J., Kushalnagar, R., Froehlich, J.: Towards accessible conversations in a mobile context for people who are deaf and hard of hearing. In: Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2018, pp. 81–92. Association for Computing Machinery, New York (2018). <https://doi.org/10.1145/3234695.3236362>
6. Kawas, S., Karalis, G., Wen, T., Ladner, R.E.: Improving real-time captioning experiences for deaf and hard of hearing students. In: Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2016, pp. 15–23. Association for Computing Machinery, New York (2016). <https://doi.org/10.1145/2982142.2982164>

7. Kheir, R., Way, T.: Inclusion of deaf students in computer science classes using real-time speech transcription. In: Proceedings of the 12th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education, ITiCSE 2007, pp. 261–265. Association for Computing Machinery, New York (2007). <https://doi.org/10.1145/1268784.1268860>
8. Kosmas, P., et al.: Enhancing accessibility in cultural heritage environments: considerations for social computing. *Univ. Access Inf. Soc.* **19**(2), 471–482 (2019). <https://doi.org/10.1007/s10209-019-00651-4>
9. Kushalnagar, R.S., Lasecki, W.S., Bigham, J.P.: Accessibility evaluation of classroom captions. *ACM Trans. Access. Comput.* **5**(3) (2014). <https://doi.org/10.1145/2543578>
10. Loizides, F., Basson, S., Kanevsky, D., Prilepova, O., Savla, S., Zaraysky, S.: Breaking boundaries with live transcribe: expanding use cases beyond standard captioning scenarios. In: The 22nd International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2020, Association for Computing Machinery, New York (2020). <https://doi.org/10.1145/3373625.3417300>
11. Majewski, J., Bunch, L.: The expanding definition of diversity: accessibility and disability culture issues in museum exhibitions. *Curator: Mus. J.* **41**(3), 153–160 (1998). <https://doi.org/10.1111/j.2151-6952.1998.tb00829.x>
12. Matthews, T., Carter, S., Pai, C., Fong, J., Mankoff, J.: Scribe4Me: evaluating a mobile sound transcription tool for the deaf. In: Dourish, P., Friday, A. (eds.) *UbiComp 2006*. LNCS, vol. 4206, pp. 159–176. Springer, Heidelberg (2006). https://doi.org/10.1007/11853565_10
13. Namatame, M., Kitamura, M., Wakatsuki, D., Kobayashi, M., Miyagi, M., Kato, N.: Can exhibit-explanations in sign language contribute to the accessibility of aquariums? In: Stephanidis, C. (ed.) *HCII 2019*. CCIS, vol. 1032, pp. 289–294. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23522-2_37
14. Namatame, M., Kitamura, M., Iwasaki, S.: The science communication tour with a sign language interpreter (2020)
15. Okuyama, K., et al.: 12.3-inch highly transparent LCD by scattering mode with direct edge light and field sequential color driving method. In: *SID Symposium Digest of Technical Papers*. Wiley Online Library (2021)
16. Olwal, A., et al.: Wearable subtitles: augmenting spoken communication with lightweight eyewear for all-day captioning. In: Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, UIST 2020, pp. 1108–1120. Association for Computing Machinery, New York (2020). <https://doi.org/10.1145/3379337.3415817>
17. Peng, Y.H., et al.: Speechbubbles: enhancing captioning experiences for deaf and hard-of-hearing people in group conversations. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018, pp. 1–10. Association for Computing Machinery, New York (2018). <https://doi.org/10.1145/3173574.3173867>
18. Proctor, N.: Providing deaf and hard-of-hearing visitors with on-demand, independent access to museum information and interpretation through handheld computers. In: Proceedings of Museums and the Web (2005)
19. Seita, M., Andrew, S., Huenerfauth, M.: Deaf and hard-of-hearing users' preferences for hearing speakers' behavior during technology-mediated in-person and remote conversations. In: Proceedings of the 18th International Web for All Conference, W4A 2021. Association for Computing Machinery, New York (2021). <https://doi.org/10.1145/3430263.3452430>

20. Seita, M., Huenerfauth, M.: Deaf individuals' views on speaking behaviors of hearing peers when using an automatic captioning app. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, CHI EA 2020, pp. 1–8. Association for Computing Machinery, New York (2020). <https://doi.org/10.1145/3334480.3383083>
21. Wald, M., Bain, K.: Universal access to communication and learning: the role of automatic speech recognition. *Univ. Access Inf. Soc.* **6**(4), 435–447 (2008)
22. White, S.: Audiowiz: nearly real-time audio transcriptions. In: Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2010, pp. 307–308. Association for Computing Machinery, New York (2010). <https://doi.org/10.1145/1878803.1878885>
23. Yamamoto, K., Suzuki, I., Shitara, A., Ochiai, Y.: See-through captions: real-time captioning on transparent display for deaf and hard-of-hearing people. In: The 23rd International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2021. Association for Computing Machinery, New York (2021). <https://doi.org/10.1145/3441852.3476551>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

