See-Through Captions: Real-Time Captioning on Transparent Display for Deaf and Hard-of-Hearing People

Kenta Yamamoto*
kenta.yam@digitalnature.slis.tsukuba.ac.jp
University of Tsukuba
Tsukuba, Japan

Ippei Suzuki*
1heisuzuki@digitalnature.slis.tsukuba.ac.jp
University of Tsukuba
Tsukuba, Japan

Akihisa Shitara*
University of Tsukuba
Tsukuba, Japan

Yoichi Ochiai
University of Tsukuba
Tsukuba, Japan

Figure 1: Proposed real-time captioning system.

ABSTRACT

Real-time captioning is a useful technique for deaf and hard-of-hearing (DHH) people to talk to hearing people. With the improvement in device performance and the accuracy of automatic speech recognition (ASR), real-time captioning is becoming an important tool for helping DHH people in their daily lives. To realize higher-quality communication and overcome the limitations of mobile and augmented-reality devices, real-time captioning that can be used comfortably while maintaining nonverbal communication and preventing incorrect recognition is required. Therefore, we propose a real-time captioning system that uses a transparent display. In this system, the captions are presented on both sides of the display to address the problem of incorrect ASR results, and the highly transparent display makes it possible to see both the body language and the captions.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility design and evaluation methods**; **Accessibility technologies**.

KEYWORDS

Real-Time Captioning, Deaf and Hard-of-Hearing, Transparent Display, Accessibility

ACM Reference Format:
Kenta Yamamoto, Ippei Suzuki, Akihisa Shitara, and Yoichi Ochiai. 2021. 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 '21), October 18–22, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3441852.3476551

*Three authors contributed equally to this research.

ASSETS '21, October 18–22, 2021, Virtual Event, USA

Table 1: Comparison with previous methods of real-time captioning.

	Mobile Device	AR Device	Transparent Display
Confirmation of ASR result by hearing person	Yes	No	Yes
Seeing body language of hearing person	No	Yes	Yes
Situation of communication	Multiple People One-to-One	Multiple People One-to-One	One-to-One

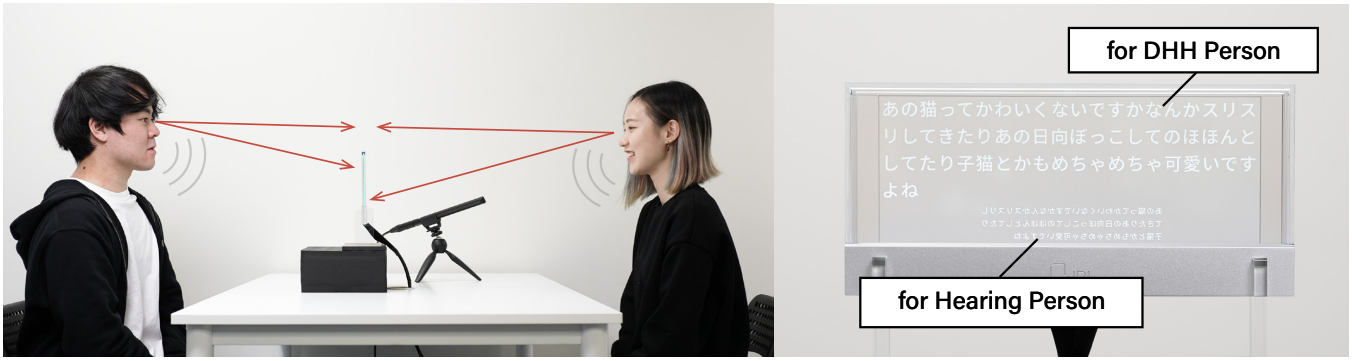

Figure 2: (left) Photograph of DHH person and hearing person using the proposed system. They talk using their voice while seeing the captions displayed on the transparent display. (right) Captions presented on the display. A small caption with left and right flips is displayed for the hearing person.

1 INTRODUCTION

Deaf and hard of hearing (DHH) people need to be more accessible to audio information in their daily lives. The importance of speech language in interpersonal communication, education, and business situations has necessitated its visualization. Visualization methods include real-time translation by an interpreter using sign language [15] and real-time speech-language transcription by a supporter [10]; however, they are expensive and thus infeasible for casual daily use. In recent years, real-time captioning via automatic speech recognition (ASR) has attracted attention because of the advances in the processing performance of mobile devices, internet-communication speed, and speech-recognition accuracy. This technology enables the provision of support to DHH people in various situations.

ASR has long been expected to serve as a universal method of accessing speech information [16]. Significant effort has been devoted to introduce such a technology in the field of education [1] and implement the use of this technology by DHH people [7–9]. The main method used in the educational setting is the capturing of teachers' speech information via ASR. After studies in a static context such as lectures in a classroom, research on the free utilization of speech recognition by DHH people in various situations has been conducted. In methods adopted in previous studies, the sound information to be recognized by DHH people was sent to translators via the internet, and the text that they verbalized was sent to mobile devices [11, 17]. With the development of high-performance mobile devices, such as smartphones and smartwatches, research on text conversion via ASR on mobile devices has increased [6]. Furthermore, with the progress made in the augmented reality (AR) technology, research on the use of AR to realize caption displays is being actively conducted [5, 6, 13, 14].

The method of realizing real-time captioning via ASR using such mobile and AR devices has some limitations. For example, when using a mobile device, the body language of the partner cannot be confirmed because the mobile device must be viewed to see the ASR result; when using an AR device, DHH people can see the speech-recognized text while observing the partner's body language. Previous studies have favored this advantage of using an AR device [4]. However, when an AR device is used, the speaker who does not wear it cannot confirm whether the speech has been correctly recognized, which may lead to errors in communication.

In this study, to address these problems, we developed a real-time captioning system that utilizes a transparent display and allows people to check the speech-recognition results while seeing their partner during the conversation (Fig. 1). Table 1 summarizes the characteristics of the proposed system and existing methods (mobile devices and AR devices). Although the proposed system is limited to one-on-one communication, we expect it to help in improving the quality of communication because it can prevent incorrect ASR and overlooking of the body language of the partner. In addition, the installation of this system in common places where voice conversation is expected, such as cash registers in supermarkets and reception desks at government offices, can possibly help DHH people in their daily lives.

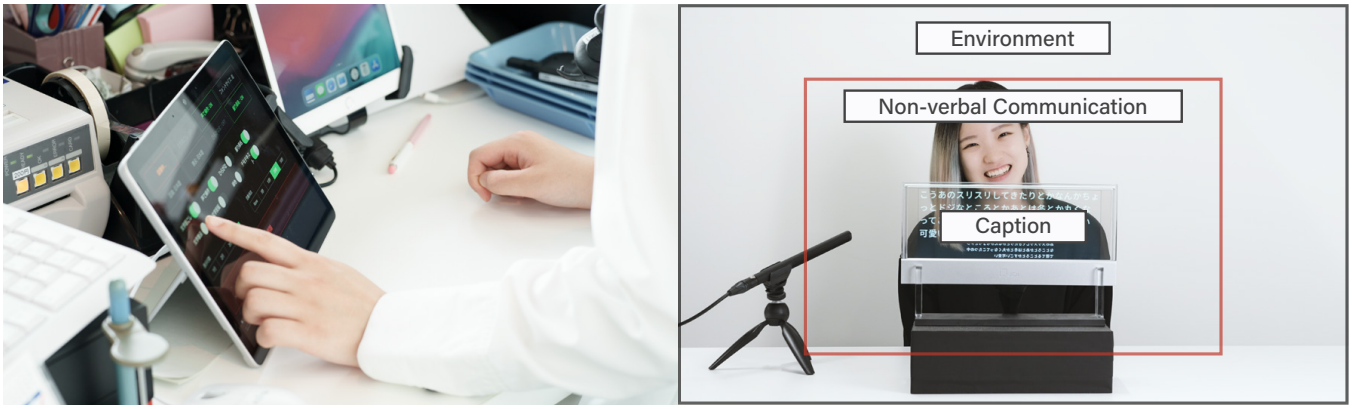

Figure 3: (left) User interface for changing the caption design. (right) Three factors for superior real-time captioning system with the transparent display. This picture was taken from the viewpoint of a DHH person when using this system.

2 PROPOSED SYSTEM

2.1 System Configuration

The proposed system presents the result of real-time captioning via ASR on a transparent display, and hearing and DHH people can communicate while checking the speech-recognition result. The two main functions of this system are ASR and caption display. First, a directional microphone is used as a speech-input device for ASR, and only the speech of the hearing person is used as the input speech. The input speech is converted into text via ASR. The system uses SpeechRecognition of the Web Speech API from Google Chrome (ver 87.0.4280.88) for ASR. Next, the text of the speech-recognition result is presented on the transparent display in real time. In this prototype, a transparent display from Japan Display Inc. is used [12], which allows for the displaying of captions on both sides of a single display, and the entire processing from ASR to caption display is performed on the web browser.

2.2 Communication Process

The proposed system is expected to be used in one-to-one and face-to-face communication. The conversation between two people across a transparent display is illustrated in Fig. 2. A hearing person speaks into a speech-input device, and the recognized speech is converted into text and presented on a transparent display. The captions are primarily meant for the DHH person to read. However, to enable the hearing person to confirm that there is no significant difference between their speech and the ASR result, the same text as the caption for the DHH person is displayed with the left and right reversed. If the ASR result is incorrect, the hearing person indicates through gestures that there was a recognition error and then repeats their speech such that it is correctly recognized. The DHH person understands the speech-language content by reading the captions on the transparent display and responding through their voice. The hearing person listens to the voice of the DHH person and inputs the voice to the microphone again. The conversation continues with a repetition of these exchanges.

2.3 Caption Design

The caption design has a significant effect on readability. Previous studies have investigated the caption-appearance preferences [2] and designs based on the reliability of speech recognition [3]. The proposed system provides an interface that allows the user to change the caption-design parameters such as character size, character color, character transparency, font, position, line spacing (Fig. 3 (left)). However, there are some design parameters that have not been examined. Therefore, by conducting a user study, it is necessary to identify the important elements in the design of captions to be displayed on a transparent display and improve the system.

3 DISCUSSION

We developed the proposed system on the premise that DHH people can speak with voice. The main reason for this is that the proposed system was developed with Shitara, who is a Deaf person and one of the authors of this manuscript, and he was able to speak. However, among DHH people, each person has his or her own preference for communication methods. Especially some people are unable to speak well and can only use the sign language. Therefore, it may not be the best to apply this system to conversations with such people. To make the communication with these people possible, efforts such as enabling automatic recognition of sign language as input from DHH people may be necessary.

In addition, the design space of this system is important. Optimal captioning requires three elements: correctly assisting in understanding of speech language, not interfering nonverbal communication, and not spoiling interior design (Fig. 3 (right)). Further studies on such design spaces will be required in the future together with user studies.

4 CONCLUSION AND FUTURE WORK

We have developed and demonstrated a see-through type real-time captioning system that utilizes a transparent display. To achieve a better system design, more user studies are required. For example, it is necessary to further clarify the merits and demerits of this system from the user studies and carefully compare them with those of mobile and AR devices. Furthermore, it is necessary to investigate