Considering Multi-Modal Speech Visualization for Deaf and Hard of Hearing People

Yusuke Toba*, Hiroyasu Horiuchi*, Shinsuke Matsumoto*, Sachio Saiki*, Masahide Nakamura*,
Tomohito Uchino†, Tomohiro Yokoyama†, and Yasuhiro Takebayashi†
*Graduate School of System Informatics, Kobe University
1-1 Rokkodai, Nada, Kobe, Hyogo 657-8501, Japan
Email: toba@ws.cs.kobe-u.ac.jp, {shinsuke, masa-n}@cs.kobe-u.ac.jp
†Special Needs Education School for the Deaf, University of Tsukuba
2-2-1 Konodai, Ichikawa, Chiba 272-8560, Japan
Email: uchino.tomohito.fu@un.tsukuba.ac.jp

Abstract—**Supporting deaf and hard of hearing (D/HH) people to understand natural conversation is one of the important activities of social welfare. However, currently the communication support for D/HH people is not enough in Japan. Although existing communication methods, such as sign language and lip-reading, are effective in one-to-one conversation, they have several disadvantages in one-to-many conversation such as meetings or conventions. In order to support D/HH people in understanding conversation, this paper proposes a multi-modal visualization application which provides many aspects of information about speech contents. Concrete examples of visualization modes include displaying subtitles by voice recognition and showing speaker's mouth to assist lip-reading.**

Keywords*—*Multi-modal visualization, deaf and hard of hearing people, supporting understanding conversation, lip-reading, voice recognition

I. Introduction

Supporting deaf and hard of hearing people to understand natural conversation is one of the important activities of social welfare. In Japan, some educational institutions provide voluntary students who write summary notes in order to help understand the contents of lessons for D/HH students. Furthermore, fair opportunity to participate in society for D/HH people achieves steady progress by revising a law about handicapped person's employment. Therefore, communication support for D/HH people is highly needed to be addressed not only by public services but also by private companies and organizations [1] [2] [3] [4].

However, currently the communication support for D/HH people is not enough [5]. One reason is that hearing impairment is difficult to judge from appearances. This feature also hinders understanding the D/HH's disability from hearing people. Some D/HH employees lost opportunity for promotion in their companies since they could not fully understand contents of seminars or meetings due to lack of supporting system. In order to improve social welfare of D/HH people, it is necessary to provide information support in which D/HH people can communication among a group of hearing people [3] [4].

Deaf people have many kinds of communication methods. Most of these methods make up for disabled hearing ability with eyesight. The most well-known method is a sign language which enables users to communicate by visual language using hands, fingers, arms and face expressions. The sign language has an advantage in communication speed which likes same as voice conversation. Making conversation by writing is also an effective approach in some high-literate countries. Some D/HH people, especially in congenitally D/HH people, gain a lip-reading skill. This skill helps understanding utterance contents from lip movements of a speaker.

Although these methods are effective in one-to-one conversation, they have several disadvantages in one-to-many conversation such as meetings or conventions. Sign language requires much effort to achieve it. Furthermore, the diffusion rate of the sign language is significantly low in hearing people. Japanese ministry has reported that even for D/HH people, only 14.1% can use sign language. Writing communication is not suitable among multiple hearing people, and takes up a lot of time and work compared with oral communication. Lip-reading is not effective in a situation that not all speakers can always show their lips to the lip-reader while they are talking. Unknown or domain-specific technical terms also difficult to guess by lip-reading.

The goal of this paper is to support understanding speech contents of hearing people for D/HH people. Especially, we focus on real-time communication with multiple hearing people such as meetings or conventions. In order to accomplish this goal, we propose a multi-modal visualization application which provides many aspects of information about contents of utterance. Here, the term *multi-modal visualization* means providing informative contents by presenting multiple visualization mode. Speech to text (STT) is one of the example of the modes. STT mode displays utterance contents as subtitles by using a voice recognition engine. Another example is a lip-reading support mode which uses a camera device and face detection engine to show enlarged movie around mouths of a speaker. We suppose that selecting and combining each visualization mode by D/HH people themselves would be helpful for understanding of conversation.

II. Scope and Challenge

In this paper, we try to support for D/HH people to understand contents of speech of hearing people as one of information support ways for D/HH people. Especially, the focus situation is real-time communication with multiple hearing people like meeting in company. Although some communication approaches (i.e., sign language, writing and lip-reading) can complement each other, it is still difficult to deal with a situation among multiple hearing people. Sign language has a critical issue in terms of its adoption rate and learning cost. Making conversation by writing in a meeting causes reduction of efficiency of communication and requires high understanding and cooperation for D/HH people from all participants.

Because of lack of ways of information support, many D/HH people have difficulty to attend meeting. As a result, they lost opportunities of promotion in their companies. Meanwhile, United Kingdom ensures enrich D/HH support as social welfare. In that country, D/HH people have fair opportunities to be assigned to important position if s/he has enough skills. Considering this fact, this loss of chance in Japan can be regarded as a high social barrier for D/HH workers. In order to provide welfare for D/HH people, it is important to provide information support way in communication among plural hearing people like meeting.

III. Multi-Modal Speech Visualization Application

A. Requirements

The purpose of this application is for D/HH people to assist understanding natural conversation by hearing people. In order to achieve the purpose, we propose an application satisfying following four requirements. These requirements are extracted from dozens years' experiences in Special Needs Education School for the Deaf, University of Tsukuba, Japan.

R1: Saving initial costs. First of all, it is important to save initial cost for both D/HH and hearing people. The initial cost includes preparing a special device, developing custom software and training particular skill.

R2: Providing various information for understanding. Utterance contents can be picked up from various perspectives of information visualization. For example, displaying subtitles by voice recognition basically helps for every D/HH person. Displaying around mouth supplements several errors of the voice recognition for D/HH people who achieves a lip-reading skill.

R3: Looking back information about speech. It is difficult for lip-reading to deal with a one-to-many situation since multiple hearing people sometimes speak at the same time. So it is useful D/HH people can look back information, such as voice or lip images, after they failed to understand part of speech.

R4: Selecting information by D/HH people. Although providing various information is an important requirement as described in **R2**, there are some cases where visualizing too much information tends to hinder understanding. Effective approaches of information visualization strongly depend on each individual D/HH person. Our system has to ensure selection of visualization modes by users themselves..

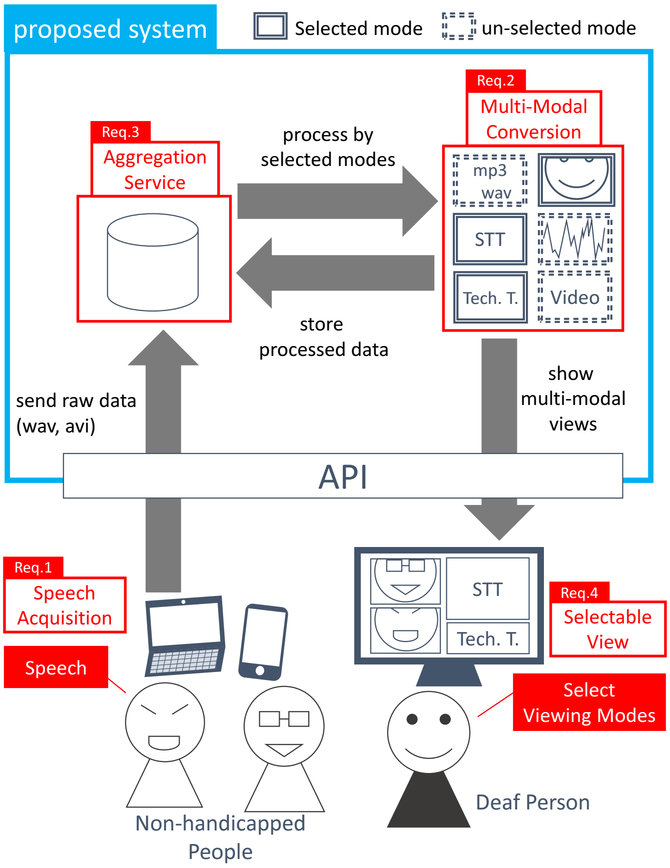


Fig. 1. Architecture of proposed application

B. Architecture

Figure 1 shows an architecture of the proposed application. Four requirements specified in Section III-A are also illustrated in this figure. At first, two hearing people use their laptop or smartphone as a client of the system. Their voice and face image data are stored in a database via such devices through APIs of the system. Stored data are processed by visualization modes which are previously selected by the D/HH user. This figure includes various examples of visualization modes such as *STT (Speech to Text) mode* and *Tech. T. (Technical Term) mode*. Details of these modes are described in the next section. Finally, the processed and visualized information are provided to the D/HH user.

C. Features

In this section, we explain four features of our proposed application based on an example usage situation. The situation is illustrated in Figure 2. In this case, four hearing users and one D/HH user participate in a meeting. They use their own devices as a client of the system.

F1: Available on various devices. This system is deployed as a web application. This application can be used from everywhere via Web browser on smartphone or personal computer as shown in Figure 2 since this is web application.

F2: Providing various multi-modal visualization. This system provides many aspects of visualization modes for

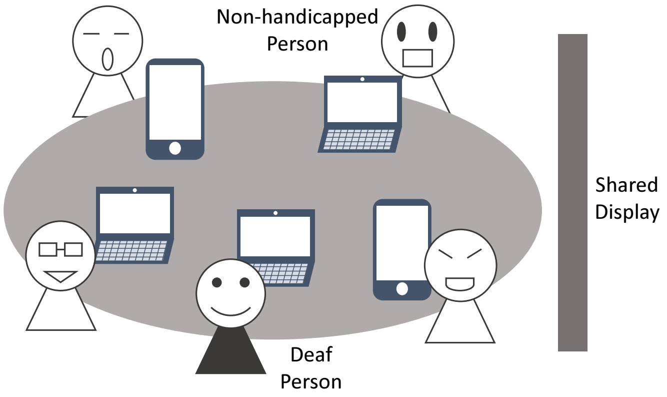
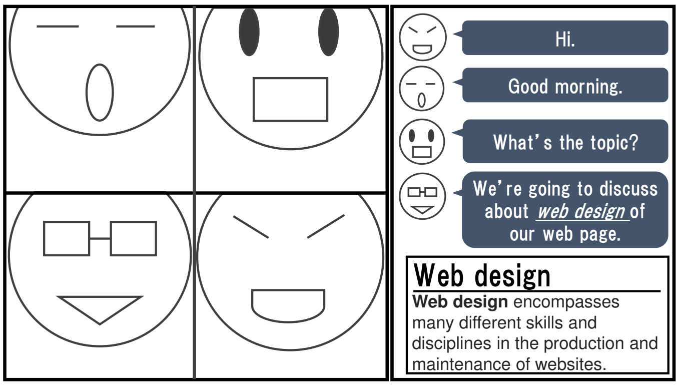


Fig. 2. An example situation of proposed application


Hi.

Good morning.

What' s the topic?

We' re going to discuss about *web design* of our web page.

Web design

Web design encompasses many different skills and disciplines in the production and maintenance of websites.

Fig. 3. An example screen of proposed application

D/HH people to understand contents of speech of hearing people. Figure 3 represents an example of screen of the system. The screen includes the following three visualization modes.

- **Speech to text mode:** This mode displays speech as subtitles by using voice recognition engine. An example of this mode is illustrated on the upper right in Figure 3.
- **Lip detection mode:** This mode magnifies and displays faces around mouth of a speaker for assist lip-reading. Faces on the left side of Figure 3 are shown by the mode. The mode will be helpful in a situation where a D/HH user is difficult to directly show a mouth of speaker due to their seat location.
- **Technical term mode:** Basically, unknown or unfamiliar word is hard to understand by lip-reading. This mode shows brief explanation of technical terms, which are possibly unknown and even quite complex, with using a domain-specific dictionary of the meeting. An example of usage is displayed on the lower right in Figure 3.

F3: Storing information about conversation. Both raw data (e.g., voice and face movies) and processed data (e.g., subtitles and detected lip movies) are stored in a database. User can look back information of conversation in past when s/he forgot or could not understand the speech at once.

F4: Selectable modes. User can select which information they use to grasp conversation from various visualization modes described in **F2**.

IV. Conclusion and Future Work

This paper proposes an application for multi-modal speech visualization, which provides many aspects of information about contents of speech to support D/HH people in a one-to-many situation.

Introducing a feature of promoting cooperation with hearing people may our proposed system more efficient. In our experiences, the accuracy of voice recognition service is not sufficient to meet actual situation of conversation. So, motivating hearing people to be cooperative is reasonable approach. An example is introducing a concept of gamification which encourages participants by showing a rank from frequency of giving and receiving spell correction in misrecognized words. This feature also prompt hearing people naturally to correct misrecognition of their speech and to speak smoothly in short sentences in order to be recognized accurately. By applying these ideas, the system can support not only understanding of speech for D/HH people, but also *comprehension of disability by hearing people*.

One of future works is combining our system with see-through head-mounted display and Augmented Reality (AR). In the system, a user has to often gaze at a display. So, when a speaker uses whiteboard in the meeting, the user has to look at the speaker, the whiteboard and the display one after another. The problem will be solved by introducing see-through head-mounted display and AR. Although it is difficult for one head-mounted display to get voice of multiple people individually, providing information not in display but in users' sight directly will let users understand conversation naturally.

Acknowledgment

This research was partially supported by the Japan Ministry of Education, Science, Sports, and Culture [Grant-in-Aid for Scientific Research (C) (No.24500079, No.24500258), Scientific Research (B) (No.26280115), Young Scientists (B) (No.26730155)] and Kawanishi Memorial ShinMaywa Education Foundation.

References

[1] M. Marschark, G. Leigh, P. Sapere, D. Burnham, C. Convertino, M. Stinson, H. Knoors, M. P. Vervloed, and W. Noble, "Benefits of sign language interpreting and text alternatives for deaf students' classroom learning," *Journal of Deaf Studies and Deaf Education*, vol. 11, no. 4, pp. 421–437, 2006.

[2] A. M. Piper and J. D. Hollan, "Supporting medical conversations between deaf and hearing individuals with tabletop displays," in *Conference on Computer Supported Cooperative Work*, 2008, pp. 147–156.

[3] R. Punch, P. A. Creed, and M. B. Hyde, "Career barriers perceived by hard-of-hearing adolescents: Implications for practice from a mixed-methods study," *Journal of Deaf Studies and Deaf Education*, vol. 11, no. 2, pp. 224–237, 2006.

[4] R. Punch, M. Hyde, and D. Power, "Career and workplace experiences of australian university graduates who are deaf or hard of hearing," *Journal of Deaf Studies and Deaf Education*, vol. 12, no. 4, pp. 504–517, 2007.

[5] A. Weisel and R. G. Cinamon, "Hearing, deaf, and hard-of-hearing israeli adolescents' evaluations of deaf men and deaf women's occupational competence," *Journal of Deaf Studies and Deaf Education*, vol. 10, no. 4, pp. 376–389, 2005.