INTERSPEECH 2017: Show & Tell Contribution
August 20–24, 2017, Stockholm, Sweden

Remote articulation test system based on WebRTC

Ikuyo Masuda-Katsuse

Kindai University, Japan
katsuse@fuk.kindai.ac.jp

Abstract

A remote articulation test system with multimedia communication has been developed in order that outside speech-language-hearing therapists (STs) can exam pronunciations of the students in special education classes in regular elementary schools and give advice to their teachers. The proposed system has video and voice communication and image transmission functions based on WebRTC. Using image transmission, the ST presents picture cards for the word test to the student and asks what is depicted. Using video / voice communication, the ST confirms the student's voice and articulation movement. Compared to our previous system in which written words were presented, the proposed system enables a more formal and accurate articulation test.

Index Terms: Articulation test, Telecare, Remote system, Speech disorders, Web application

1. Introduction

In Japan, special education classes in regular elementary schools support children whose language development lags behind their peers due to hearing / speaking disorders or developmental disabilities. Since such experts as speech-language-hearing therapists (STs) are not assigned to classes and school teachers generally lack formal training in speech education techniques, the teachers have to learn such techniques on their own or cooperate with outside experts. The teachers who are greatly motivated to improve their own teaching methods sometimes consult the outside experts. However, it is difficult for the outside experts to give advice suitable for individual students to the teachers because they rarely directly examine students' pronunciations.

From such a background, we developed a support system for pronunciation instruction / practice in special education classes for language-disabled children to promote cooperation between teachers and STs [1]. In this system, while students practiced their pronunciation using the system, the speech sounds for articulation tests were automatically recorded on a web server. The outside STs could access them through the Internet anytime and anywhere and perform articulation tests. The STs could then use the results in collaboration with the teachers [2].

On the other hand, in the previous system, the outside STs had to perform articulation tests using only speech sounds. Referencing students' mouth shapes during utterance may aid in more accurate diagnosis for lateralized articulation and so on. Therefore, in this study, a remote articulation test system with multimedia communication has been developed.

At INTERSPEECH 2017, the prototype system will be exhibited. Remote articulation tests can be simulated with two tablet PCs connected to the Internet for outside STs and students respectively.

2. Articulation test in Japan

The articulation test authorized in Japan [3] consists of several tests as follows. The word test is the main one, and the others are optional.

(1) Word test: Articulation is tested for a fixed set of 50 words. In the test, subjects are presented with licensed picture cards and asked what they depict.

(2) Observation of conversation: Speech intelligibility is evaluated by observation of free-talk.

(3) Syllable test: Subjects are asked to repeat syllable sounds uttered by a tester.

(4) Sound test: Whether mispronunciation changes or not by presenting sound stimuli is tested.

(5) Sentence test: Subjects are asked to repeat sentences uttered by a tester.

(6) Similar movement test: Basic movement for articulation or similar movement is tested.

3. System configuration

3.1. Outline

The proposed system has video and voice communication and image transmission functions based on WebRTC [4]. The communication is performed between the outside ST and the special education class. Using image transmission, the ST presents picture cards for the word test to the student. Using video / voice communication, the outside ST confirms the student's voice and articulation movement. Compared to the presentation of written words in the previous system [1], this system is able to present picture cards, so it becomes possible to test more formally. In addition, most of the optional tests (2) – (6) can also be performed using video / voice communication.

3.2. Entering a web conference room

In this system, in order to achieve communication between the special education class and the outside ST, the "web conference room" model was adapted. The teacher (or the student) and the outside ST input their IDs and the room name on a website to enter the conference room. The partners who can communicate are limited to those in the same room. The outside ST selects his / her partner's ID and establishes a session with him / her.

3.3. Peer to peer session establishment

To directly transmit and receive video / audio data and image files between the student and the outside ST, RTC peer Connection of WebRTC is used. To establish peer connection between them, it is necessary to exchange some information

about their session, such as Offer / Answer message of SDP [5]. A signaling server is required to mediate such information. In this research, the server was developed using socket.IO [6], which is a library of Node.js [7].

Furthermore, in order to communicate beyond Network Address Translation (NAT), it is necessary to acquire the global IP address and port number allocated by NAT. In this research, this information is obtained using Google's publicly available STUN [8] server.

The outside ST clicks the “Call Start” button on the website of the web conference, and a signal for communication start is sent to the student's computer. Then, direct communication is performed by sending Offer /Answer of SDP or ICE Candidate [9] to each other. In this research, trickle ICE [10] was adopted to exchange the ICE Candidate.

3.4. Media acquisition and image transmission

Camera images and microphone sounds are acquired using getUserMedia of WebRTC. The data are sent by SDP and presented on the browser.

Furthermore, if necessary, the outside ST can transmit images such as picture cards, syllables, or sentence cards necessary for articulation examination to the student. After the ST presses the "Browse" button on the screen, the ST can select the image file on a file explorer. After selecting the image file and pressing the "Send Image" button, the image is converted to blob formatted binary data and transmitted to the student side using DataChannel of WebRTC, and then is displayed on the browser on the student's side. A thumbnail of the image is also displayed on the ST’s browser.

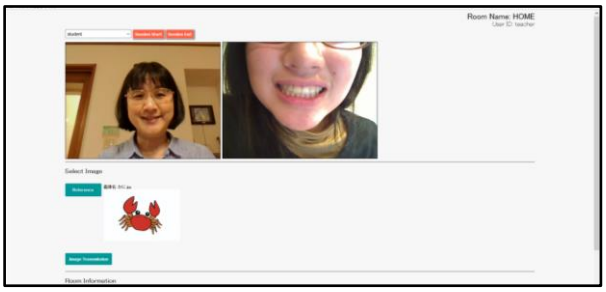

Figure 1: *Web screen of STs. Video of the ST is displayed on the upper left side and the student video is displayed to right of that. A thumbnail of the image transmitted by the ST is displayed below the video window.*

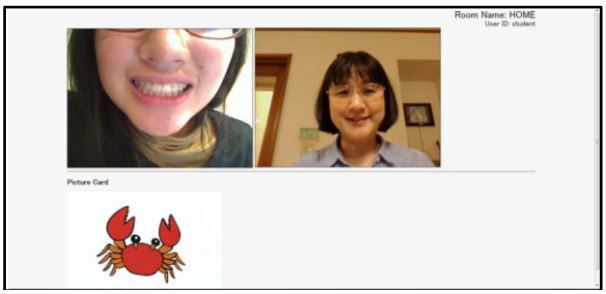

Figure 2: *Web screen of students. Video of the student is displayed on the upper left side and the ST video is displayed to right of that. The image transmitted by the ST is displayed below the video window.*

4. Discussion and future works

In this paper, we proposed a remote articulation test system using video / voice communication and image transmission function based on WebRTC. The ability to conduct remote inspection will reduce the burden on outside STs, teachers, and students, and promote collaboration between them. However, the accuracy and ease of conducting the test when using the proposed system are required to be equal or better compared with a face-to-face articulation test. Matthews et al. [11] investigated the acceptability of using Skype to provide speech and language therapy (SLP). Their results showed that a Skype SLT session was acceptable. We also plan to conduct a demonstration experiment in order to confirm the usefulness of the proposed system.

Currently, we are expanding the proposed system. One example of expansion is that user management in the proposed system will be linked with the previous system [1]. As another example, by adding other simple functions to the proposed system, it is also possible to check for some dyslexia problems.

In Japan, officials are considering whether to apply a similar medical-fee system to telemedicine and telecare as that for face-to-face medical treatment. In the future, the demand for telemedicine / telecare systems with functions matching the objectives of individual medical care and education, such as the proposed system, will increase.

5. Acknowledgements

Takuya Yoshida, an undergraduate student, contributed to system development. Part of this work was supported by JSPS KAKENHI Grants 16K00496.

6. References

[1] I. Masuda-Katsuse, “Web application system for pronunciation practice by children with disabilities and to support cooperation of teachers and medical workers,” *Proceedings of INTERSPEECH 2015,* pp. 1066–1067, 2015.

[2] I. Masuda-Katsuse, “Web application system to support pronunciation instruction in special education classes for language-disabled children,” *Proceedings of YKJCA 2017,* pp. J04-1–J04-4, 2017.

[3] Articulation Clinics Workshop [Ed], *Japanese Articulation Test, Second Edition*, Chiba Test Center, 2010. *(in Japanese)*

[4] A. B. Johnston, D. C. Burnett, *WebRTC: APIs and RTCWEB Protocols of the HTML5 Real-Time Web*, Digital Codex LLC, 2014.

[5] J. Rosenberg, and H. Schulzrinne, “An Offer/Answer Model with the Session Description Protocol (SDP)”, *RFC3264*, 2002.

[6] R. Rai, *Socket.IO Real-Time Web Application Development*, Packt Publishing, 2013.

[7] Node.js : https://nodejs.org/

[8] J. Rosenberg, R. Mahy, P. Matthews, and D. Wing, “Session Traversal Utilities For NAT (STUN)”, *RFC 5389*, 2008.

[9] J. Rosenberg, “Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols”, *RFC5245*, 2010.

[10] E. Ivov, E. Rescorla, and J. Uberti. "Trickle ICE: Incremental Provisioning of Candidates for the Interactive Connectivity Establishment (ICE) protocol." *IETF Internet Draft, draft-ietf-ice-trickle-04*, 2016.

[11] R. A. Matthews, B. Woll, and M. Clarke, “Researching the acceptability of using Skype to provide Speech and Language Therapy,” *International Journal of Integrated Care,* 12(4), 2012.