ULTRASOUND-ENHANCED MULTIMODAL APPROACHES TO PRONUNCIATION TEACHING AND LEARNING

Jennifer Abel, Blake Allen, Strang Burton, Misuzu Kazama, Bosung Kim, Masaki Noguchi, Asami Tsuda, Noriko Yamane, and Bryan Gick
University of British Columbia.

1 Introduction

Pronunciation is an integral part of communication, as it directly affects speakers' communicative competence and performance, and ultimately their self-confidence and social interaction. Second language (L2) pronunciation is one of the most challenging skills to master for adult learners. Explicit pronunciation instruction from language instructors is often unavailable due to limited class time; even when time is available, instructors often lack knowledge of effective pronunciation teaching and learning methods. Imitating native speakers' utterances can be done outside of the classroom, but the absence of feedback makes it difficult for learners to improve their skills.

In an effort to improve pronunciation instruction, the Department of Linguistics and the Japanese language program in the Department of Asian Studies at the University of British Columbia began a collaboration in 2014 to develop new multimodal approaches to pronunciation teaching and learning. Japanese is the largest language program at UBC, with more than 1,500 students enrolled every year, and is known to be the most diverse in terms of learners' language backgrounds. The project has developed online resources to allow learners of Japanese to improve their pronunciation, as well as to allow Linguistics students to better understand sound production. The project's key technological innovation is the use of ultrasound overlay videos, which combine mid-sagittal ultrasound images of tongue movement in speech with external profile views of a speaker's head to allow learners to visualize speech production [1]. A strong focus to this point has been on the pronunciation of Japanese sounds, with instruction incorporating explicit awareness of tongue movements and insights from articulatory phonology, including the interference of L1 phonology on L2 learning [3].

2 Methods

Ultrasound of native speakers of Japanese and of English was recorded using an Aloka ProSound SSD-5000 system, and the exterior video was recorded using a JVC camcorder (GZ-E300AU). Both recordings were made at 30 frames per second. The exterior video showed the left profile of the speaker's head. A clapper was used to generate an audio alignment point.

The ultrasound overlay videos were created from raw footage using a four-step process. First, the ultrasound and exterior video were trimmed using Adobe Premiere to ensure alignment. Next, all elements of the ultrasound image aside from the tongue were manually erased using Adobe After Effects. The brightness of the tongue was increased, and the colour was changed from white to a shade of pink (colour #DE8887 in Adobe After Effects) to more closely resemble the human tongue. Then, the erased ultrasound image was overlaid on the exterior face video using Adobe After Effects. Scaling of the two sources was achieved by ensuring that the shadow of the probe in the ultrasound image is the same width as the top of the probe in the exterior video. The results of this process are exemplified in Figure 1.

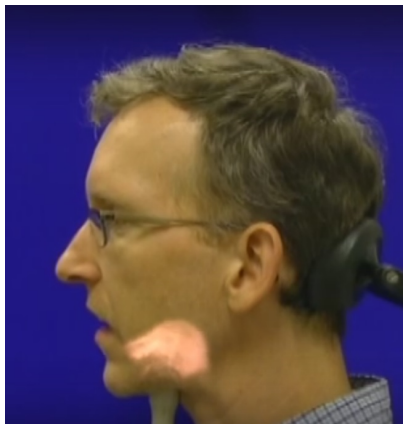

Figure 1. Ultrasound overlay video frame of [χ]

3 Results

The videos are available to the public through the eNunciate website (http://enunciate.arts.ubc.ca/), and are licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. The videos are categorized into 'Linguistics' and 'Japanese' content, although all pages are open to all students and instructors.

3.1 Linguistics content

The Linguistics pages feature ultrasound overlay videos of canonical examples of the sounds of the world's languages produced basic contexts: in the [_a] and [a_a] positions for consonants, and in isolation or in [C_C] contexts for vowels. Freeze frames are inserted in these videos to capture key moments in the articulation (e.g., the stop closure in a stop articulation), and beginning and end titles are inserted. These videos can be accessed through interactive IPA consonant and vowel charts. In addition to the ultrasound overlay videos, videos introducing the use of ultrasound in linguistics and the basics of vowel and consonant articulation are available. To date, these videos have been used mainly in introductory Linguistics courses as in-class

demonstration materials or as optional out-of-class resources.

3.2 Japanese content

The Japanese pages include instructional and exercise videos for Japanese pronunciation teaching and learning. These videos incorporate narration, cartoons, and animations in addition to ultrasound overlay elements, and are augmented with quizzes to allow students to reinforce what they have learned using the videos. The videos are grouped into three categories: introductory, which includes introductions to Japanese sounds and to basic phonetic concepts; 'challenging sounds', which features videos focusing on problems that L2 learners from different language backgrounds may encounter; and intonation.

These videos have been used in both introductory and upper-year Japanese classes as pronunciation teaching and assessment tools. In the introductory classes, students are expected to watch particular 'challenging sounds' or intonation videos as preparation for in-class quizzes. In the upper-year class in which the materials were used (JAPN 301 at UBC), the students were encouraged to use the videos to improve their fluency when reading aloud. Instructors and students have anecdotally reported satisfaction with the resources; full-scale effectiveness and satisfaction data collection is in progress.

4 Discussion: developments in progress

4.1 Biofeedback tools

As part of our plan to use biofeedback to facilitate L2 pronunciation learning, we are developing two tools that will allow learners to visualize their speech in real-time. The first tool is an interactive tongue visualizer, which will automate creation of the type of ultrasound overlay videos described in section 2 based on ultrasound and video feeds of a speaker producing sounds in real time. Development of this tool is still in the early stages. The visualizer will be implemented at a physical location ("Pronunciation Station") at UBC, and will be equipped with a CHISON ECO 1 portable ultrasound with a 6.0MHz D6C12L Transvaginal probe.

The second tool is a web app that allows visualization of speech pitch contours based on open-source audio manipulation software [4]. Students will be able to visually compare the pitch of their own pronunciations of L2 speech with pitch contours produced by native speakers, allowing them to improve their L2 intonation, pitch accent, and/or tone patterns. Students will also be able to save their pronunciations in order to track how their pronunciation improves over time. Currently, we have implemented a basic web interface for the app, along with an implementation of the standard autocorrelation algorithm for inferring pitch from sound signals. Our immediate goals are 1) to refine the pitch detection algorithm to allow it to better compensate for the rapid pitch variations in natural speech, and 2) to build a more user-friendly interface.

4.2 Ultrasound training

To overcome the lack of a standardized procedure for the teaching of L2 pronunciation with ultrasound imaging, we are developing guidelines based on the procedures previously used in the settings of L2 learning [1] and speech language pathology [2]. The guidelines target three consecutive days of teaching to allow teachers to use the Pronunciation Station: (1) initial evaluation of students' pronunciation, (2) training with ultrasound images as biovisual feedback, and (3) post-training evaluation of students' pronunciation. As a case study, we will implement the protocols in teaching Japanese pronunciation to native speakers of Korean, particularly focusing on the acquisition of the contrast between alveolar and alveo-palatal sibilants (e.g. [za] vs. [ʑa]), which is known to be especially difficult for Korean speakers. The effectiveness of the protocols will be discussed in light of the data collected at the Pronunciation Station.

5 Conclusion

As more research continues to support the positive effects of multimodal feedback via technologies such as ultrasound on L2 pronunciation training [e.g. 5], the importance of bringing these technologies into the classroom has become clear. We believe that the tools and techniques developed as part of the eNunciate project are helping to achieve that goal at UBC, and hope to extend them to other institutions and situations in the future.

Acknowledgments

This project is supported by a Flexible Learning Large Project Grant from the Teaching and Learning Enhancement Fund at the University of British Columbia. Many thanks to Joe D'Aquisto, Jonathan de Vries, Amir Entezaralmahdi, Lewis Haas, Tsuyoshi Hamanaka, Hisako Hayashi, Ross King, Andrea Lau, Yoshitaka Matsubara, Douglas Pulleyblank, Nicholas Romero, Hotze Rullmann, Murray Schellenberg, Joyce Tull, Martina Wiltschko, Jenny Wong, and Kazuhiro Yonemoto.

References

[1] Gick, B., et al. (2008). Ultrasound imaging applications in second language acquisition. In J. G. Hansen Edwards and M. L. Zampini (eds.), *Phonology and Second Language Acquisition* (pp. 309-322). Amsterdam: John Benjamins.

[2] Bernhardt, B., et al. (2005). Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics & Phonetics* 19.6-7: 605-617.

[3] Toda, T. (2003). *Second Language Speech Perception and Production: Acquisition of Phonological Contrasts in Japanese*. Lanham, Maryland: University Press of America.

[4] Cwilso/PitchDetect. (n.d.). Retrieved August 11, 2015 from https://github.com/cwilso/PitchDetect.

[5] Pillot-Loiseau, C., et al. (2015). French /y/-/u/ contrast in Japanese learners with/without ultrasound feedback: vowels, non-words and words. Paper presented at ICPhS 2015. Retrieved August 12, 2015 from http://www.icphs2015.info/pdfs/Papers/ICPHS0485.pdf.