Using ultrasound for teaching and researching articulation

Ian Wilson*

Center for Language Research, Department of Computer Science and Engineering, The University of Aizu, Tsuruga, Ikki-machi, Aizu-Wakamatsu, 965–8580 Japan

Keywords: Ultrasound, Articulation, Speech research, Pronunciation teaching

PACS number: 43.70.Aj, 43.70.Jt, 43.70.Kv [doi:10.1250/ast.35.285]

1. INTRODUCTION

Safely viewing and accurately describing the motion of the tongue during speech has been a challenging problem for many years [1]. Although methods of viewing the tongue during speech have existed for decades, many of these methods are not ideal for various reasons: too expensive (MRI), too dangerous (X-ray, CT scan), too troublesome (electromagnetic articulometry — EMA), or too limited in the data one can get (electropalatography — EPG). Ultrasound, however, is becoming cheaper, is safe, is easy to set up and use, and is able to provide real-time images of the whole tongue during speech. The continuing development of diagnostic ultrasound machines as tools to view the tongue during speech has contributed to both speech research and pronunciation teaching. An excellent overview to using ultrasound for speech research can be found in [2].

Ultrasound was used for speech research as long ago as the late 1960s [3], but in early studies, either the ultrasound was a one-dimensional beam (like a laser pointer), which gave information only along that one dimension, or the field of view of the probe was so narrow that it did not provide a clear view of the whole tongue. Now, almost 50 years later, three-dimensional ultrasound machines exist, although the most commonly used machines for pronunciation teaching and speech research are two-dimensional ones — it is often difficult to interpret and make measurements of three-dimensional images that are displayed on two-dimensional screens. The field of view of most probes used for speech research today ranges from 90° to over 150°, enabling a view of the whole tongue — from the tongue tip to the tongue root.

Diagnostic ultrasound’s high-frequency (3–16 MHz) sound waves go through skin, fat and muscle easily, but they get absorbed by bone and reflect off air boundaries. For the best view of the tongue, then, the probe must be held against the neck, below the chin, so that the sound waves penetrate the tongue from below (see Fig. 1). With a narrow enough probe and a little bit of upward pressure against the neck, the sound waves freely travel between the mandible and the hyoid bone, through the muscle of the tongue, reflecting back off the air above the surface of the tongue. Depending on the orientation of the probe, a midsagittal or a coronal image can be obtained. Clinical research using coronal tongue imaging by ultrasound is described in [4]. A typical midsagittal tongue image obtained using ultrasound is shown in Fig. 2. Note that in Fig. 2, the tongue tip is on the right.

As mentioned above, ultrasound is easy to set up and use. By simply holding the probe to the neck with one’s hand, it is possible to get a clear view of the tongue’s shape and movements during speech — sufficient for a qualitative assessment of the tongue’s movements. In most pronunciation teaching settings, this is all that is required. However, if one needs more precision, as in the case of collecting data to make quantitative measurements for research, then care must be taken not to have movement of the probe relative to the head. Even in controlled data collection settings with the probe held on a metal stand and the back of the head leaning against a wall, the head still moves an average of 1–4 mm in the front-back dimension and 4–9 mm in the up-down dimension [5]. If the head and probe move relative to each other, the image on the screen moves too, and so it appears that the tongue is moving. Thus, to achieve the highest precision, such probe movement must be controlled or must be corrected for. It can be controlled by mechanically preventing the probe and head from moving, and it can be corrected for by tracking the movement during data collection and then later computationally factoring out that movement with software. In such a short overview article, it is impossible to go into detail about all of the methods that exist for controlling and correcting for head movement relative to the probe, but some methods are described in [6–8]. As for how to analyse the ultrasound data after collecting it, a multitude

*e-mail: wilson@u-aizu.ac.jp
of methods exists for quantitative tongue shape analysis, and an excellent comparison of some of these methods can be found in [9].

Ultrasound machines that are used in hospitals are usually very large (cart-mounted) and costly. Many speech laboratories and speech language pathologists’ offices have neither the space nor funding to purchase such machines. Fortunately, though, many portable models of ultrasound machines exist and are used in fieldwork outside the lab or clinician’s office [10]. There are even USB ultrasound probes, which simply plug into one’s own laptop computer. At Ultrafest VI in November 2013, Alan Wrench conducted a survey of attendees asking them what ultrasound systems they were currently using. Ultrafest is a conference held approximately once every 2 years and is specifically for speech researchers and educators who use ultrasound in their work. The results of Wrench’s survey, which were presented by him at the conference [11], can be seen in Table 1.

For recording ultrasound movies of the tongue’s motion during speech, one needs to combine the video signal output from the ultrasound machine with an audio signal from a microphone that the participant speaks into. Depending on the hardware and software used, synchronization of these two signals has been known to be problematic.

Regarding the video signal, although the machine-internal frame rate is usually higher than 100 frames per
second, the screen on the machine is limited to standard video resolution: 30 frames per second in Japan and North America, and only 24 frames per second in Europe. Depending on the type of output and the settings one uses, a higher frame rate may be obtained (see [12], for example).

Since ultrasound makes the whole tongue visible, it is especially good for research involving the tongue root, because no other safe and cost-effective imaging method enables one to see the tongue root. An example of such a study is [13], using ultrasound to give clear evidence of the mapping between a phonological feature and its articulatory gesture. Another type of research that benefits from seeing the whole tongue at once is investigations of inter-articulator timing patterns (see [14] for one example). Researchers in such studies are trying to determine the relative timing of speech gestures made by different parts of the tongue (i.e., tongue tip, tongue dorsum, tongue root).

2. USING ULTRASOUND IN SECOND-LANGUAGE PRONUNCIATION CLASSROOMS

There exist two general approaches in the teaching and learning of foreign-language pronunciation today: (a) Intuitive-Imitative approaches and (b) Analytic-Linguistic approaches [15]. The former approach traditionally involves imitating the teacher’s voice without any explicit phonetic information, whereas the latter approach makes use of tools such as the International Phonetics Alphabet (IPA), articulatory charts and information, and also physical models such as in [16]. A recent study [17] has shown the effectiveness of giving explicit phonetic information (i.e., articulatory configurations) to Japanese learners of English pronunciation.

As the use of Computer-Assisted Language Learning (CALL) software became more common, the acoustic signal could be viewed and compared to a pronunciation model. However, the mapping between acoustics and articulation is often difficult to interpret, so the learner does not know how to proceed. Systems that infer the learner’s articulation from the acoustic signal exist but they have the danger of drawing false inferences and incorrectly judging the learner’s pronunciation as right when it is wrong, or as wrong when it is right (see [18], though, for successful ultrasound tests of such a system). Systems that directly measure a learner’s articulation and provide biofeedback are less likely to make such mistakes.

For the phonetically naive language learner, movies displaying a teacher’s correct tongue movements during speech are much easier to interpret than spectrograms. The same is true about biofeedback of one’s own tongue moving in real-time on the screen. So, since at least as early as 2006, ultrasound has been used in some university classrooms for teaching and learning pronunciation [19–21]. From January to March 2009 (and repeated again in 2010), the CLR Phonetics Lab at the University of Aizu was regularly featured on a weekly NHK program, bringing ultrasound images of the tongue to Japanese television every week.

There are a number of English sounds for which ultrasound imaging can enlighten the Japanese learner. One pair of sounds is North American English /l/ and /r/, which do not occur in Japanese. The initial consonant sound in Japanese /ra/, /ri/, /ru/, /re/, and /ro/ is typically produced using only the tongue tip. On the other hand, English /l/ and /r/ are articulatorily complex sounds — /l/ has two gestures (tongue tip raising and tongue dorsum retraction), while /r/ has three gestures (tongue tip or tongue blade raising, tongue root retraction, and lip rounding). When students actually see the difference between these two English sounds, as well as the difference between those and the Japanese sound, there is a greater chance of them being able to produce the sounds. In fact, one may hypothesize that a learner should be able to produce each sound (with the help of ultrasound feedback) even if she or he is not able to perceive the differences. In an ultrasound study of native Japanese speakers, it was recently shown that the perception of English /r/ and /l/ undergoes a different developmental course from the production of those sounds [22].

Another set of sounds for which ultrasound can help the Japanese learner of English is the tense-lax vowel pairs, for example /i/ and /ɪ/ (“heat” versus “hit”), or /u/ and /ʊ/ (“Luke” versus “look”). Not only is the tongue body slightly lower in the lax vowels, but the tongue root is not advanced forward like it is for the tense vowels.

If one turns the probe 90° for a coronal view of the tongue, it is possible to see midline tongue grooving for /s/, but the opposite — tongue doming for /ʃ/. In this way, students can see when they are mistakenly pronouncing “seat” like “sheet,” for example, and what they must do to fix the problem.

Showing the pronunciation learner his or her own tongue while teaching articulation is obviously difficult to do with a large class and only one ultrasound machine. However, the advantage of such real-time biofeedback is clear when one considers that learners are not always able to correctly follow a teacher’s verbal instructions on where and how much to move the tongue. In [6], an ultrasound study was done asking eight participants to move their tongues a certain direction. Participants were not shown the ultrasound screen, but simply had to follow the movement instructions that were given. Results showed that participants were surprisingly poor at (blindly) moving their tongues as instructed. However, when participants are shown their own tongue on the screen, it is relatively
straightforward for them to follow articulatory instructions by a teacher. In the future, it is anticipated that the cost of ultrasound USB probes will come down far enough to make it feasible to have many probes available in a large-scale pronunciation classroom.

Articulatory setting, the underlying language-specific setting of the tongue, lips, jaw, etc. during speech, is another area of foreign-language pronunciation teaching where research has been made possible through the use of ultrasound. In [23], ultrasound and other data has shown that bilinguals who sound native-like in both of their languages actually have different rest positions for each language during speech. This finding has important implications for the way that foreign-language pronunciation is taught in the future.

Another area where ultrasound can enhance pronunciation instruction is with hearing-impaired learners and others who are undergoing speech therapy. Some speech-language pathologists make extensive use of ultrasound to diagnose patients’ speech problems and to help correct these problems together with the patients (see [24] for a good overview, and also [25–27]).

It is clear that the number of speech researchers, clinicians, and foreign-language pronunciation teachers using ultrasound has increased greatly over the last decade. As a safe, easy-to-use method of viewing the whole tongue in real-time during speech, ultrasound is sure to attract more and more users as the cost comes down in the near future.

REFERENCES

[1] B. Gick, I. Wilson and D. Derrick, Articulatory Phonetics (Wiley-Blackwell, Chichester, UK, 2013).
[2] M. Stone, “A guide to analysing tongue motion from ultrasound images,” Clin. Linguist. Phonot., 19, 455–501 (2005).
[3] C. A. Kelsey, F. D. Minifie and T. J. Hixon, “Applications of ultrasound in speech research,” J. Speech Hear. Res., 12, 564–575 (1969).
[4] M. Kato, Y. Ohtsuka, Y. Mukai, S. Imatomi, T. Kimura and F. Shusse, “Tongue movements during lateral articulation: Observation by standardized ultrasound analysis in coronal section,” Jpn. J. Logop. Phoniatr., 43, 30–39 (2002).
[5] N. Horiguchi and I. Wilson, “Head movement and tongue compression with a fixed ultrasound probe but a free head,” Ultrafest 4, New York University (2007).
[6] N. Horiguchi, “How L2 pronunciation learners interpret articulation instructions: An ultrasound study of the tongue,” MSc Thesis, University of Aizu (2012).
[7] T. Bressman, H. Flowers, W. Wong and J. C. Irish, “Coronal view ultrasound imaging of movement in different segments of the tongue during paced recital: Findings from four normal speakers and a speaker with partial glossectomy,” Clin. Linguist. Phonot., 24, 589–601 (2010).
[8] J. Mielke, A. Baker, D. Archangeli and S. Racy, “Palatron: A technique for aligning ultrasound images of the tongue and palate,” Coyote Papers: Working Papers in Linguistics, Linguistic Theory at the University of Arizona, 14, 96–107 (2005).
[9] K. M. Dawson, D. H. Whalen and M. K. Tiede, “A comparison of methods for tongue shape analysis,” Ultrafest 6, University of Edinburgh (2013).
[10] B. Gick, S. Bird and I. Wilson, “Techniques for field application of lingual ultrasound imaging,” Clin. Linguist. Phonot., 19, 503–514 (2005).
[11] A. Wrench, “Ultrasound speech analysis: State of the art,” Ultrafest 6, Edinburgh, Scotland (2013), http://materials.articulateinstruments.com/Technical/State_of_Art.ppt (accessed July 12, 2014).
[12] A. L. Miller and K. B. Finch, “Corrected high-frame rate anchored ultrasound with software alignment,” J. Speech Lang. Hear. Res., 54, 471–486 (2011).
[13] F. Hudu, “[ATR] feature involves a distinct tongue root articulation: Evidence from ultrasound imaging,” Lingua, 143, 36–51 (2014).
[14] F. Campbell, B. Gick, I. Wilson and E. Vatikiotis-Bateson, “Spatial and temporal properties of gestures in North American English /r/,” Lang. Speech, 53, 49–69 (2010).
[15] M. Celce-Murcia, D. M. Brinton, J. M. Goodwin and B. Griner, Teaching Pronunciation, 2nd ed. (Cambridge University Press, New York, 2010).
[16] T. Ara, “Mechanical models of the human vocal tract,” Acoust. Today, 9, 25–30 (2013).
[17] K. Saito, “Reexamining effects of form-focused instruction on L2 pronunciation development,” Stud. Second Lang. Acq., 35, 1–29 (2013).
[18] O. Engwall, “Analysis of and feedback on phonetic features in pronunciation training with a virtual teacher,” Comput. Assist. Lang. Learn., 25, 37–64 (2012).
[19] B. Gick, B. M. Bernhardt, P. Bacsfalvi and I. Wilson, “Ultrasound imaging applications in second language acquisition,” in Phonology and Second Language Acquisition, J. G. H. Edwards and M. L. Zampini, Eds. (John Benjamins, Amsterdam, 2008), pp. 309–322.
[20] H. M.-L. Tsui, “Ultrasound speech training for Japanese adults learning English as a second language,” MSc Thesis, University of British Columbia (2012).
[21] B. Meadows, “Do you see what I’m saying?: Ultrasound technology as a tool for pronunciation instruction,” in Communication Technology for Students in Special Education and Gifted Programs, J. E. Aitken, J. P. Fairley and J. K. Carlson, Eds. (Information Science Reference, Hershey, Pa, 2012), pp. 293–302.
[22] M. Tateishi, “Effects of the use of ultrasound in production training on the perception of English /r/ and /l/ by native Japanese speakers,” MA Thesis, University of Calgary (2013).
[23] I. Wilson and B. Gick, “Bilinguals use language-specific articulatory settings,” J. Speech Lang. Hear. Res., 57, 361–373 (2014).
[24] B. Bernhardt, B. Gick, P. Bacsfalvi and M. Adler-Bock, “Ultrasound in speech therapy with adolescents and adults,” Clin. Linguist. Phonot., 19, 605–617 (2005).
[25] M. Adler-Bock, B. Bernhardt, B. Gick and P. Bacsfalvi, “The use of ultrasound in remediation of English /r/ in two adolescents,” Am. J. Speech-Lang. Pathol., 16, 128–139 (2007).
[26] H. B. Klein, T. McAllister Byun, L. Davidson and M. I. Grigos, “A multidimensional investigation of children’s /r/ productions: Perceptual, ultrasound, and acoustic measures,” Am. J. Speech-Lang. Pathol., 22, 540–553 (2013).
[27] J. L. Preston, N. Brick and N. Landi, “Ultrasound biofeedback treatment for persisting childhood apraxia of speech,” Am. J. Speech-Lang. Pathol., 22, 627–643 (2013).
Ian Wilson is a professor in the Center for Language Research at the University of Aizu. He received his PhD in Linguistics from the University of British Columbia in 2006. He teaches a graduate course called “Speech Articulation and Acoustics,” as well as undergraduate courses in Pronunciation, Thesis Writing, and EFL. He was a regular on NHK Television’s program in 2009 and 2010, and he is co-author of “Articulatory Phonetics” (Wiley-Blackwell Publishers, 2013).