



# Ultrasound biofeedback in pronunciation teaching and learning

S. Bird<sup>1</sup>, B. Gick<sup>2</sup>

<sup>1</sup>University of Victoria, Canada, <sup>2</sup>University of British Columbia, Canada,  
<sup>1</sup>*sbird@uvic.ca*, <sup>2</sup>*gick@mail.ubc.ca*

## Abstract

Ultrasound imaging is an important and increasingly accessible tool for visualizing the speech articulators in pronunciation teaching and learning. In this paper, we start with an overview of the importance of pronunciation in language learning and teaching, and of speech visualization as a tool for supporting pronunciation work in this context. We then briefly introduce ultrasound imaging as a speech visualization tool and review the history of ultrasound biofeedback in speech research, from clinical studies to second language studies. We follow this with an introduction to some of the available ultrasound systems and their pros and cons and to relevant methodological considerations. We end with a summary of the value of speech visualization in the context of pronunciation teaching and learning.

**Keywords:** Technology assisted language learning (TALL), speech production, speech visualization, ultrasound imaging.

## 1. Introduction

Although recent work in second language (L2) pronunciation has focused on intelligibility and comprehensibility [1] rather than on ‘correct accent’, pronunciation is an important aspect of L2 acquisition. Often, it is the first indication of a speaker’s linguistic abilities, from the perspective of a listener [2]; in some cases, e.g. in the context of Indigenous language revitalization, intelligibility and comprehensibility are not enough: learners are keenly aware that they are responsible for transmitting their language to future generations, and are very committed to pronouncing it in a way that honours their elders [3]. Recent work has also shown that accurate pronunciation bootstraps other aspects of language learning, such as word learning [4].

While pronunciation is an important aspect of L2 learning, it has often been challenging to teach and learn in traditional classroom settings, especially in cases where the target sounds are not

easily visible and/or where awareness of how sounds are being articulated is lacking. For example, Salish languages have a velar/uvular contrast that carries a significant functional load and is therefore important to master; however, understanding the articulatory difference between /k kw/ and /q qw/ is challenging for learners [5, 3], in part because we as speakers (of any language) tend to have poor awareness of what our speech articulators are doing in the back of our mouths. This is where technology-assisted language learning can help. Indeed, learning pronunciation is a multimodal activity, and visualizing speech benefits learners (e.g., [6, 7, 8]).

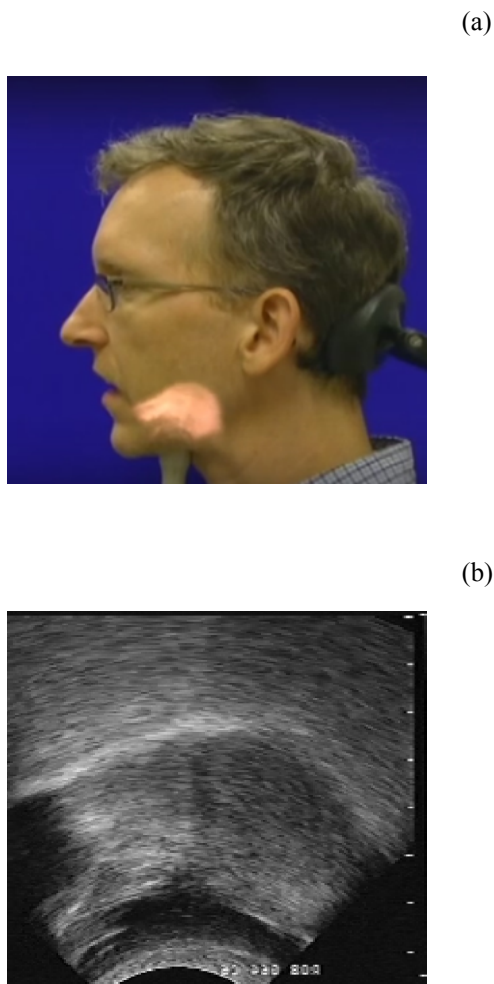
The aim of this paper is to provide an introduction to ultrasound imaging in particular, as a biofeedback tool for teaching and learning pronunciation.

## 2. Ultrasound as a biofeedback tool

A growing literature on visual feedback has evaluated the benefits to L2 learners of being able to see the moving speech articulators during speech; these studies have used a range of tools, including ultrasound, electromagnetic articulography (EMA) and electropalatography (EPG) [9]. Of these, ultrasound is particularly safe, non-invasive, and increasingly affordable and portable - including to the classroom, where the display can be projected through standard audio-visual technology. Ultrasound also has the advantage of having been created as a visualization tool used across a wide range of fields, making it an ideal tool for providing visual biofeedback during speech: it is user-friendly and widely available, and its output (dynamic videos of tongue movement) is very intuitive.

To image the tongue during speech, the transducer (also known as a “probe”) is stabilized under the chin (see Figure 1a). The transducer sends out ultra-high frequency sound waves (hence “ultrasound”), which are reflected back when they

hit the air just above the tongue surface. The time it takes the sound waves to travel to the tongue surface and back to the transducer is converted to a set of distance estimates between the probe and the tongue surface; the tongue contour tracing (see Figure 1b) is a visual composite of these distance estimates.



**Figure 1:** Ultrasound imaging set-up (a) and generated tongue contour, with the tip on the right (b).

### 3. Brief history of lingual ultrasound pronunciation research

Lingual ultrasound research began several decades ago, with pioneering work such as [10]. Since the early 2000s, intervention studies have established the effectiveness of ultrasound biofeedback with clinical populations such as hard of hearing (e.g., [11, 12]). In subsequent studies, ultrasound was used to teach individual challenging sounds, such as English /r/, in clinical settings (e.g., [13]). The

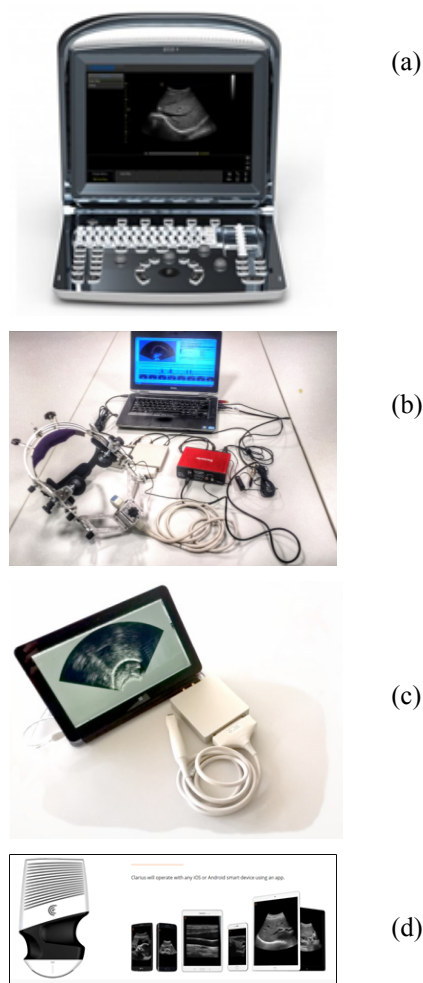
earliest L2 pronunciation studies, starting with [14], used methods that were based in this earlier clinical tradition, finding dramatic results with small numbers of individually trained participants. While all of these early studies shared the same methodology of one-on-one instruction in a formal laboratory setting, more recent L2 pronunciation research is showing that ultrasound can also be adapted for use in larger classroom settings, either by providing individual ultrasound training to language instructors [15] or by presenting ultrasound videos as part of a blended learning approach [16]. Our own experiences with ultrasound technology have shown that it is also successful in community-based settings, with mixed groups (in terms of proficiency, age, and comfort level with technology) that differ from typical language classroom demographics.

Although we see great potential for incorporating ultrasound into the teaching and learning of pronunciation, very few studies exist on ultrasound-enhanced pronunciation teaching and learning in any context. Existing studies tend to be focused on small numbers of learners (ten or fewer) acquiring small numbers of segments (e.g., English /r, l/). No studies that we are aware of have systematically compared ultrasound imaging with other approaches to teaching pronunciation or provided measures of student engagement or long-term outcomes. With increased accessibility of ultrasound systems, however, we hope to see an expansion of research assessing the applications of ultrasound imaging in pronunciation intervention.

### 4. Available systems

A great deal has changed since the early days of lingual ultrasound applications, both in terms of the types of systems available and their associated costs. Generally, size and speed (including temporal and spatial resolution) correlate positively with price. This relationship is the opposite of that in technology more generally, where smaller equals better equals more expensive. By way of example, two of the machines housed in the UBC Interdisciplinary Speech Research Lab (ISRL) as of this writing are: an older, full-sized Aloka SSD-550 and a newer, semi-portable Chison Eco 1 (Figure

2a). The Aloka is extremely powerful, with built-in functionality allowing for precise quantitative research on the machine itself. The trade-off is that it is not portable, and it cost approximately \$150,000 CAD over a decade ago. Similar-sized machines today still bear a similarly high price tag. The Chison, in contrast, is less powerful, with fewer special features, and is well suited for qualitative studies and for teaching applications. It is portable, and costs approximately \$5,700 USD. This can play into the hands of speech researchers and pronunciation teachers, for whom portability can be a major advantage.



**Figure 2:** Examples of ultrasound systems currently in use for speech research and teaching: (a) Chison Echo 1; (b) Articulate Instruments' research system; (c) Articulate Instruments' clinical (SonoSpeech) system; (d) Clarius wireless probe.

Portable systems currently come in two varieties: the Chison is a complete, self-contained system, including a built-in monitor and control

panel, transducer, and self-contained processing and visualization software. Increasingly, however, manufacturers are making systems that run on a user's own external computer or handheld device. Articulate Instruments offers two packages specifically designed for speech, one geared towards quantitative research (Figure 2b), the other towards clinical intervention (Figure 2c). The research system comes with a transducer, head stabilizer, and software for experimental research, including precise time alignment of audio and ultrasound streams (frame rate for the ultrasound stream is 60-120 Hz). The whole system sells for under 7000 GBP (as of late 2018). The SonoSpeech (clinical) system is much simpler, without the components required for quantitative research (head stabilization and audio-video synchronization). Several features of the SonoSpeech are potentially very valuable for pronunciation teaching and learning, although we have not yet experimented with these ourselves. In particular, it has overlays to mark up regions of the display that learners should focus, and template videos of target/model pronunciations can be placed side by side with the live display. The software is actively being developed, with plans to also include automatic segmentation and automatic live contour fitting, for example. It sells for around 4000GBP. Both Articulate Instruments systems require a PC or Mac running Windows and can be powered via USB; the SonoSpeech software can be used on a tablet (it is touch screen enabled). Although the two systems are marketed as distinct, it is in fact possible to "mix and match" features of both, and in our experience the manufacturer is very responsive, in terms of accommodating users' specific needs.

More and more manufacturers are now making stand-alone probes. Two examples are SeeMore (\$4900 CAD and Clarius (\$10,000 CAD; Figure 2c). SeeMore has a system specifically designed for clinical applications (speech pathology), potentially also useful for teaching. It is compatible with Windows machines, the transducer connecting via a USB port. The frame rate is relatively low (15 Hz), making the system useful for visualization but not for researching anything but the slowest tongue movements. Clarius makes a wireless probe that comes with an App compatible with a range of smart devices (iOS and Android) as well as cloud storage. Currently, the Clarius probe is, in our opinion, a bit too large to hold comfortably under a speaker's chin, but the design is likely to improve in

the coming months/years, as the company responds to user feedback.

Being the only package that is designed specifically for linguistic research, the Articulate Instruments system comes optimized for this application. The package includes a helmet to secure the probe to the chin, so that speakers can move their heads freely without changing the position of the probe relative to the chin, a useful feature for researchers wishing to compare articulatory properties of different sounds. To ensure comparability of images for ultrasound-based research using other systems, experimenters must find other ways of either (a) immobilizing the probe against the chin during data acquisition [17] or (b) correcting for head movement after data collection, re-aligning the tongue contours [18, 19] so that they can be compared. This is not to say that any one system is necessarily superior to others for all applications, but rather that different systems are suited to different needs, and with the many systems on the market today, it is now possible – and important – to find a machine that can fit to one’s own applications. The list of systems described above is by no means intended to be an exhaustive list, and the technology and the details of specific products will of course change quickly; this kind of diversity of options will likely continue to be available in the future.

## **5. Methodological considerations**

### **5.1 Pros and cons of ultrasound biofeedback**

As a tool for visualizing tongue position and movement during speech, ultrasound has several advantages: the whole tongue is visible (although the tip is sometimes obscured by the mandible); dynamic movement can be imaged as well as static position; imaging provides immediate biofeedback; data are relatively easy to collect; systems are relatively inexpensive; and the methods are non-invasive and familiar (compared to other imaging tools). Of course, no system is perfect, and ultrasound does have some drawbacks: it can only image the tongue (and, to a limited degree, the larynx [20]), so must be coupled with another tool for studying inter-articulator timing (e.g. between the tongue and lips); conducting quantitative research is relatively complex; and not all speakers

image equally well (generally, children and smaller adults tend to image best).

### **5.2 Appropriate research/teaching applications**

In terms of research/teaching foci, ultrasound is not equally well-fitted to imaging all aspects of speech, because of the considerations outlined above. Sounds that tend to be good candidates for ultrasound biofeedback are vowels, rhotic sounds (see Figure 3), retroflex, velar and uvular consonants, and dynamic movements between tongue gestures [21, 22]. Ultrasound cannot be used to examine non-lingual articulations (although limited studies do involve laryngeal ultrasound [19]), and tongue tip articulations are sometimes obscured by the mandible (depending on the speaker).

Because of its portability and focus on visualization, ultrasound can be used in a wide range of research and teaching contexts. We have experience with one-on-one sessions (e.g. an L2 speaker sitting with an L1 speaker and mimicking their articulations), small group sessions (e.g., in community settings with speakers and learners of Indigenous/minority languages), in larger classroom sessions (with the ultrasound display projected onto a large screen), and through online/self-directed learning portals (e.g., UBC’s eNunciate: <http://enunciate.arts.ubc.ca/>).

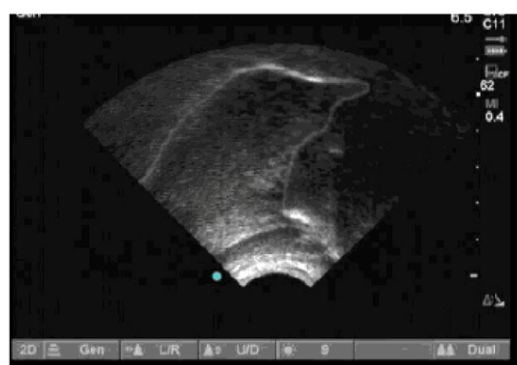
Incorporating ultrasound into pronunciation teaching does not require conducting research, and research on the benefits of ultrasound biofeedback need not necessarily involve analysis of ultrasound data per se. For example, a typical intervention study might include two groups of participants who differ only in the type of training they receive: ultrasound biofeedback vs. another tool/technique. The two groups might be exposed to similar training sequences: a pre-test, followed by a period of training, followed by a post-test (and possibly an additional subsequent delayed post-test). For such a study, the data to be analyzed might include listener judgments and/or acoustic analysis of pre- vs. post-training productions and/or participants’ self-reflections on the benefit of the training. There are, though, many ways to use ultrasound imaging for pronunciation teaching and research that range beyond the scope of the present paper.

### **5.3 Ultrasound data capture and analysis**

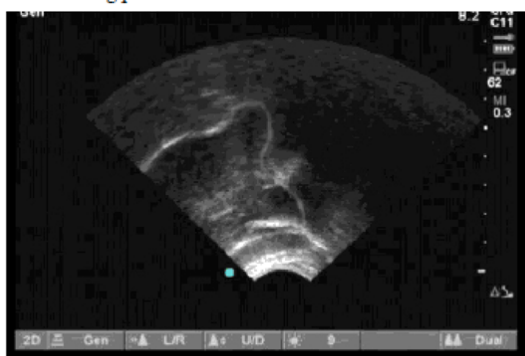
For analyzing ultrasound data specifically, there are several options with respect to data capture and



analysis. Some systems (e.g. the Aloka SSD-550 and the Chison Echo 1 mentioned above) are optimized to produce very clear images, but require an external computer or video recording device for capturing and storing videos longer than a few seconds (for subsequent analysis). Other systems, such as the Articulate Instruments package, run on a personal computer, allowing ultrasound videos to be captured and stored directly. A third storage option that is likely to become increasingly accessible is cloud storage, the method used for the Clarius wireless probe.



Pre-training production of /ɪ/ in "Read"



Post-training Production of /ɪ/ in "Read" by the same speaker

**Figure 3:** Pre- (above) and post-training (below) tongue contours for English /ɪ/ in "read", as spoken by a single Japanese (L1) English (L2) speaker. Original source (following Creative Commons license CC-BY-NC-SA 2.0): Figure 8, "Ultrasound images of one participant's improved /ɪ/ word" [23, p. 58].

Ultrasound data analysis generally involves capturing both ultrasound video and audio, and merging the two signals together, so that the audio track can help with identifying target sounds on the ultrasound track. This is not necessarily a complex process, but does require the means to synchronize the audio and ultrasound recordings. This can be done simultaneously during recording or can be part of post-processing.

The simplest type of ultrasound analysis is qualitative, and involves simply identifying the

overall shape of tongue configurations at a particular moment in time (e.g. the point of maximal constriction for a consonant or the midpoint of a vowel), based on still images extracted from the video.

Quantitative analysis requires more processing and sometimes complex methods. Quantitative analysis can be conducted on static images capturing the tongue at a particular moment in time or on videos capturing tongue movement patterns over time, for which M-mode ("movement mode") or kymography [24] can be useful. The details of quantitative data analysis are beyond the scope of this paper, but can be found in primary sources such as [25] (see [26] for a collection of relevant papers) or by comparing methods used in specific experimental studies.

As mentioned above, some types of quantitative data analysis of ultrasound images require ensuring that ultrasound videos are controlled with respect to the orientation of the probe. This can be done using head stabilization, or by post-processing after data-collection, to correct for head movement and re-align tongue contours (see above). Head and probe stabilization need not be particularly invasive (see [27]), and there are methods of quantitative analysis that do not require head stabilization at all, such as those which quantify tongue shapes rather than precise location in space (e.g., [28-31]).

## 6. Discussion and Conclusion

It is clear from our own experiences, as well as from a growing body of literature, that ultrasound can be a valuable part of the language teaching and learning toolkit. Nonetheless, many questions remain about the breadth of its usefulness and its limitations, for example: To what extent does it help with pronunciation beyond the target sound being focused on? Can it help with perception in addition to production? Can students learn passively (through simply watching ultrasound videos) as well as they can through interactive engagement (biofeedback) with the tool? Although these questions remain to be answered, we argue that ultrasound has already assumed a position as a valuable part of the language teaching and learning toolkit.

## 7. Acknowledgements

The authors wish to thank Heather Bliss for her valuable input into this work, and the audience at ISAPh2 for their questions and comments. We gratefully acknowledge support from SSHRC and from NSERC Discovery grant RGPIN-2015-05099.

## 8. References

- [1] Derwing, T., Munro, M. 2009. Putting accent in its place: Rethinking obstacles to communication. *Language Teaching* 42(4), 476–490.
- [2] Derwing, T. Rossiter, M., Murray, M., Tomson, R. 2004. Second language fluency: Judgments on different tasks. *Language Learning* 54(4), 655–679.
- [3] Bird, S., Kell, K. 2017. The role of pronunciation in SENĆOŦEN language revitalization. *Canadian Modern Language Review* 73(4), 538–569.
- [4] Johnson, K. A., Mellesmoen, G. M., Lo, R. Y-H., Gick, B. (2018). Prior pronunciation knowledge bootstraps word learning. *Frontiers in Communication: Language Science*, 3.
- [5] Bird, S. 2016. A phonetic study of the “K sounds” across generations of SENĆOŦEN speakers. UBCWPL 42: *Proceedings of the 51st International Conference on Salish and Neighbouring Languages (ICSNL)*, 19–36.
- [6] Bliss, H., Bird, S., Gick, B., Burton, S., Cooper, A. 2018. Seeing speech: Ultrasound-based multimedia resources for pronunciation learning in Indigenous languages. *Language Documentation & Conservation* 12, 315–338.
- [7] Chun, D. M. 2013. Computer-assisted pronunciation teaching. In: C. A. Chapelle (ed), *The Encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell.
- [8] Levis, J. M., Pickering, L. 2004. Teaching intonation in discourse using speech visualization technology” *System* 32(4), 505–524.
- [9] Bliss, H., Abel, J., Gick, B. 2018. Computer-assisted visual articulation feedback in L2 pronunciation instruction: A review. *Journal of Second Language Pronunciation* 4(1), 129–153.
- [10] Sonies B. C., Shawker T. H., Hall, T. E., Gerber L. H., Leighton S. B. 1981. Ultrasonic visualization of tongue motion during speech. *Journal of the Acoustical Society of America*, 70(3), 683–686.
- [11] Bernhardt, B., Gick, B., Bacsfalvi, P., Ashdown, J. 2003. Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *Clinical Linguistics & Phonetics*, 17(3), 199–217.
- [12] Bernhardt, B., Gick, B., Bacsfalvi, P., Adler-Bock, M. 2005. Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics & Phonetics*, 19(6/7), 605–617.
- [13] Adler-Bock, M. B., Bernhardt, B., Gick, B., Bacsfalvi, P. 2007. The use of ultrasound in remediation of English /t/ in two adolescents. *American Journal of Speech Language* 16(2), 128–139.
- [14] Gick, B., Bernhardt, B., Bacsfalvi, P., Wilson, I. 2008. Ultrasound imaging applications in second language acquisition. In: Hansen Edwards, J.G., Zampini, M. L. (eds), *Phonology and Second Language Acquisition*. Amsterdam: John Benjamins, pp. 309–322.
- [15] Noguchi, M., Yamane, N., Tsuda, A., Kazama, M., Kim, B., Gick, B. 2015. Towards protocols for L2 pronunciation training using ultrasound imaging. Poster presentation at the 7<sup>th</sup> annual Pronunciation in Second Language Learning and Teaching (PSLLT) Conference. Dallas, Texas, October 2015.
- [16] Bliss, H., Cheng, L., Schellenberg, M., Lam, Z., Pai, R., Gick, B. 2016. Ultrasound technology and its role in Cantonese pronunciation teaching and learning. Paper presented at the 8<sup>th</sup> annual Pronunciation in Second Language Learning and Teaching (PSLLT) Conference, University of Calgary, August 2016.
- [17] Derrick, D., Carigna, C., Chen, W.-R., Shujau, M., Best, C. 2018. Three-dimensional printable ultrasound transducer stabilization system. *Journal of the Acoustical Society of America* 144(5), EL 392–398.
- [18] Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert- Lehouillier, H., Vatikiotis-Bateson, E. and Hailey, D. S. 2005. The Haskins Optically Corrected Ultrasound System (HOCUS). *Journal of Speech, Language, and Hearing Research*, 48, 543– 553.
- [19] Mielke, J. 2015. An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *Journal of the Acoustical Society of America* 137(5), 2858–2869.
- [20] Moisik, S., Lin, H., Esling, J. 2014. A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *Journal of the International Phonetic Association* 44(1), 21–58.
- [21] Bird, S. 2012. Cool thing about ultrasound #17: Now I can pronounce /hiqəʔ/! UBCWPL: *Proceedings of the 47th International Conference on Salish and Neighbouring Languages*, 1–12.
- [22] Gick, B., Bliss, H., Michelson, K., Radanov. B. 2012. Articulation without acoustics: “Soundless” vowels in Oneida and Blackfoot. *Journal of Phonetics* 40, 46–53.
- [23] Tsui, H.M. 2012. Ultrasound speech training for Japanese adults learning English as a second language. Unpublished MSc thesis, University of British Columbia.
- [24] Liu, Y., Tong, F., Choi, D. and Gick, B. Cross-linguistic bracing: analyzing vertical tongue movement. *Canadian Acoustics*. In press.
- [25] Davidson, L. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America* 120(1), 407–415.
- [26] Stone, M. (Ed.) 2005. *Journal of Clinical Linguistics and Phonetics* 19(6–7) Special issue on ultrasound imaging of speech.

- [27] Gick, B., Bird, S., Wilson, I. 2005. Techniques for field application of lingual ultrasound imaging. *Journal of Clinical Linguistics and Phonetics* 19(6-7), 503-514.
- [28] Bressmann, T., Thind, P., Uyl, C., Bollig, C., Gilbert, R. W., Irish, J. C. 2005. Quantitative three-dimensional ultrasound analysis of tongue protrusion, grooving and symmetry: Data from 12 normal speakers and a partial glossectomee. *Clinical Linguistics and Phonetics* 19, 573-588.
- [29] Ménard, L., Aubin, J., Thibeault, M., Richard, G. 2012. Measuring tongue shapes and positions with ultrasound imaging: A validation experiment using an articulatory model. *Folia Phoniatrica et Logopaedica* 64, 64-72.
- [30] Stolar, S., Gick, B. 2013. An index for tongue curvature. *Canadian Acoustics* 41, 11-15.
- [31] Dawson, K. M., Tiede, M. K., Whalen, D. H. 2016. Methods for quantifying tongue shape and complexity using ultrasound imaging. *Clinical Linguistics and Phonetics* 30 (3-5), 328-344.