ULTRASOUND TONGUE IMAGING IN SECOND LANGUAGE LEARNING

Tanja Kocjančič Antolík

Many learners of a second language (L2) are faced with a difficulty when trying to produce new L2 segments. They often do not succeed, and the resulting speech can be judged by the native speakers as being mispronounced and having a foreign "accent". Moreover, speech sound production does not necessarily improve with the amount of L2 learning, and the difficulties can persist even when learners achieve native-like level in other aspects of L2 (Forsberg et al., 2014).

According to the two main theories of L2 speech acquisition, the main source of these pronunciation troubles lies in the inadequate perception of the new L2 speech sounds.

Speech Learning Model (Flege, 1995) states that the creation of a new phonetic category is more likely when an L2 sound is not perceptually similar to the existing native language (L1) sounds. To acquire a new L2 sound, a learner first needs to perceive a phonetic (dis)similarity between this new sound and the native speech sound system. If adequate perceptual dissimilarity is noted, a new category can be formed. If the L2 sound is perceived as similar to an existing L1 sound, it is more likely to be assimilated to a similar native segmental category. The resulting production of the new L2 sounds thus depends on the perception of the same sound, and the production can never be more adequate than the perception. Speech Learning Model also suggests that the ability to form new categories, and by extension to perceive phonetic differences, diminishes with age.

Similarly, Perceptual Assimilation Model for L2 Learners (Best & Tyler, 2007) also places the perceptual similarity between L1 and L2 sounds in the center of the L2 speech acquisition problem. This model focuses specifically on the acquisition of an L2 minimal contrast and proposes different scenarios, depending on whether none, one or both items of a minimal pair are perceived as equivalent to an L1 sound. Again, the creation of a new phonetic category is most likely when an L2 sound (or contrast) is perceived as different than any existing L1 category.

The primacy of speech perception in L2 speech sound learning is noted also in L2 classrooms. Most often it is expected that L2 learners will pick up the correct production by listening to a teacher or a model recording and trying to mimic what they hear. When specific pronunciation training is included in classroom learning, it is usually also based on perception. Although different systematic methods have been proposed to facilitate L2 speech acquisition, high variability phonetic training (HVPT) has received the most research interest (for review see Barriuso & Hayes-Herb, 2018; Thomson, 2018). HVPT, proposed first by Logan et al. (1991), employs a speech sound identification task with natural training items in varied phonetic contexts and produced by different speakers. The method has shown significant improvements not only in perception but also in the production of previously poorly acquired L2 sounds, and the retention of gains over longer periods (Bradlow, Akahane-Yamada, Pisoni & Tohkura, 1999; Lambacher et al., 2005; Wong, 2012). How-

ever, at least one study has shown no improvement in production following HVPT (Thomson & Derwing, 2016). The degree of improvement has been further linked to the number of the learner's L1 phonetics categories (Iverson & Bronwen, 2009), with more categories being beneficial, and to the learner's perceptual abilities (Perrachione, Lee, Ha & Wong, 2011). It is thus not surprising that some learners do not acquire new L2 categories adequately and manifest obvious mispronunciations even after several years of L2 learning.

Another method employed in L2 classroom learning is instruction with an explicit description of the target articulatory movements. The articulatory description has received more focus in recent years, and its positive contribution to L2 speech learning has been noted in several studies (Arteaga, 2000; Aliaga-García & Mora, 2009, Derwing, Munro & Wiebe, 1997, 1998). However, speech sound learning via articulatory description can be hindered by poor understanding of one's own articulatory movements, poor ability to describe the target movements and to follow articulatory instruction.

Derwing and Munro (2005) provide a review of several studies showing that many teachers of L2 English received no training on how to teach pronunciation. Similarly, an extensive study among 175 Australian speech and language therapists showed that even professionals most trained on articulation have inadequate awareness of tongue-palate contact in the articulation of individual speech sounds (McLeod, 2009). The majority of speech sounds (18 out of 24) were described correctly by less than 12% of speech and language therapists.

A poor ability to execute even simple instructions on tongue movements was demonstrated in an ultrasound study by Ouni (2011). The results showed that none of the 24 participants was able to successfully execute two repetitions of any of the 12 movement tasks. However, the study also demonstrated that even a short practice with ultrasound tongue imaging as a visual feedback (UTI-VF) improves the execution of the same instructions. Following a pre-test, 14 participants received a 20-minute session during which they observed their tongue movements with UTI-VF, while the control group had a 20-minute rest. At post-test, the experimental group showed improved performance at 10 out of the 12 tasks, while the control group only at three and a notable decrease at the rest.

Real-time visual feedback (VF) thus seems very appropriate for L2 speech sound learning. It makes it possible to observe the tongue during speech, to compare tongue movements of teacher and learner, to increase awareness and understanding of one's own movements, and to facilitate motor control and execution of the target movements. In terms of motor learning, it provides two types of feedback: knowledge of results (information on whether the target was successfully realized or not) and knowledge of performance (information on how the movement was realized) (Ballard et al., 2012).

Because tongue is the most active articulator involved in the production of almost all speech sounds and mostly not visible during speech, it has received the most attention in the articulatory VF research. Currently, three different methods are used for tongue visualization: electropalatography (EPG), electromagnetic articulography (EMA) and ultrasound tongue imaging (UTI).

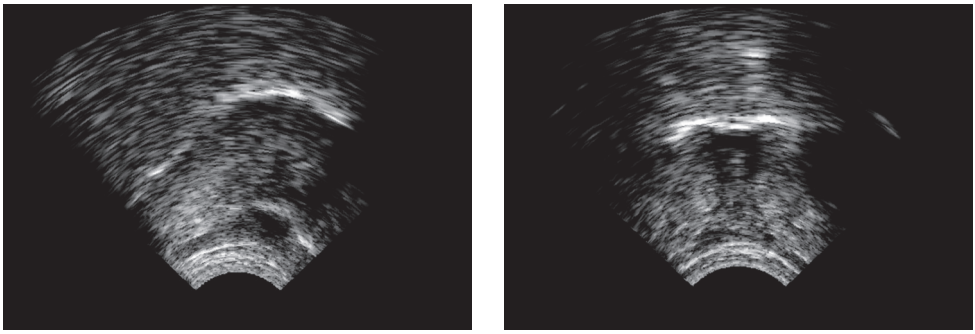

FIGURE 1: UTI, midsagittal view on the left (the front of the tongue is on the right side of the image) and coronal view on the right.

EPG images tongue-palate contact by inserting an artificial palate with embedded electrodes into the speaker's mouth. Each time the tongue touches an electrode, the contact is recorded and presented in an EPG diagram. The artificial palate covers the majority of the palate, with possibly missing out on the velar area; the artificial palate cannot extend to the very back of the oral cavity to avoid triggering a gag reflex in users. A major limitation of the EPG is that because the artificial palate has to match the user's palate as much as possible, it has to be made individually. The method is also relatively invasive since a foreign item is placed into the oral cavity and the user has to get used to it.

Even more invasive is EMA, where small electrodes are glued directly on the tongue (typically three to five electrodes) and different parts of the face (e.g., the lips). The speaker sits with their head in a method-specific magnetic field that allows tracing of the electrodes in a 3D space with high temporal and spatial resolution. In this way, it is possible to observe the movement of different parts of the tongue (front, mid, back, sides) and compare them to other articulators, such as lips.

Finally, UTI uses a medically approved ultrasound device to image the tongue surface. The method is the least invasive since it involves only placing the ultrasound probe under the speaker's chin. UTI allows observing the tongue from the midsagittal view (left image in Figure 1), from the front to the back of the tongue along the midline, or coronal view (right image in Figure 1), from one side of the tongue to the other, and it creates an image about every 15ms. The techniques allow visualization of tongue shape (bunched or flat tongue body, central groove, raised or lowered tongue sides, raised front of the tongue), position (front/back, high/low) and direction of movement during speech, as well as silent parts of articulation, such as lingual articulatory setting and any movements preceding audible speech (Wilson & Gick, 2014). Although it does not directly image any other structures inside the oral cavity, it is possible to view the hard palate if the speaker holds a small amount of water in the mouth. The main limitations of UTI are its inability to image a raised tongue tip, as well as speaker-dependent image quality.

All three methods have been successfully used in studies investigating different phonetic phenomena, as well as VF for speech sound remediation. However, because

of the greatest ease of application, UTI has become the most frequently used method for VF on tongue movements.

Currently, two commercially available systems are available for UTI-VF in clinical practice: PI 7.5 MHz Speech Language Pathology 99-5544 by Seemore (Canada) and SonoSpeech by Articulate Instruments (UK). Both systems contain an ultrasound probe than can be plugged into a Windows-running computer and come with software that allows recording and storing tongue images, as well as graphical annotation of the recorded images (e.g. to mark the palate or target tongue shape/position).

UTI-VF has been employed as a successful method for speech sound remediation in speech therapy. The method has shown faster achievement of articulatory goals, transfer to untrained items, retentions of therapy outcomes over longer periods and user satisfaction (Adler-Bock, Bernhardt, Gick & Bacsfalvi., 2007; Bernhardt, Gick, Bacsfalvi & Adler-Bock, 2005; Cleland, Scobbie & Wrench, 2015; Preston et al., 2014).

Following positive results observed in clinical speech sound remediation, the extension of the method to the L2 learning started being examined (Wilson & Gick, 2006). To date, only a handful of studies have systematically investigated using UTI-VF in L2 learning. They include research on learning new L2 speech sounds and remediation of inadequately acquired ones.

Three studies evaluating the contribution of UTI-VF in learning new speech sounds showed no advantage over the following non-VF methods: articulatory description with modeling the instructor's production (Cleland, Scobbie, Nakai & Wrench, 2015), audio recordings of target productions (Roon, Kang & Whalen, 2020) and articulatory description with static midsagittal tongue contours (Lin, Cychosz, Shen & Cibelli, 2019). The studies included a relatively large number of participants, 30 children (aged 6 to 12 years), 18 adults and 49 adults, respectively. Importantly, however, the participants were not learning an L2 language and the question remains how they would perform if the training occurred as part of an L2 course.

Contrary to the studies exploring learning new L2 sounds, those exploring L2 speech sound remediation provide greater benefit of the usage of UTI-VF, although the method was tested on a significantly smaller number of learners. Gick et al. (2008) reported an improvement in the English /ɹ — l/ contrast by three Japanese learners after only 30 minutes of training with UTI-VF. Improvement on /l/ but not on /ɹ/ was observed in ten Japanese speakers of English following five 30-minute UTI-VF sessions (Tateishi & Winters, 2013). The same method was beneficial also for three Italian speakers practicing the English /æ — ʌ/ contrast who all improved after one 30-minute session, compared to three control speakers receiving no training (Sisinni, d'Apolito, Fivela & Grimaldi, 2016). Similarly, four Japanese learners of French improved in their production of French /u/ and /y/ following three 45-minute sessions, while two learners receiving the same amount of more traditional pronunciation training with articulatory description and modelling did not show the same improvement (Kocjančič Antolík, Pillot-Loiseau & Kamiyama, 2019). Moreover, the UTI-VF group retained the gains and even showed further improvements at two months post-training. The main limitation of these studies is, however, that they all involve individual training sessions and thus a small number of participants.

Overall, the reported results on using UTI-VF in L2 learning have been encouraging. Learners were able to better understand their production and the target one, to correct their production rather quickly, and they have kept the correct production of the new sounds over time. Importantly, the learners were happy with the method and have expressed interested to use it in any future L2 learning (Kocjančič Antolík, Pillot-Loiseau & Kamiyama, 2019). However, more studies are needed to validate the method on a larger sample of learners of different languages, to track long-term outcomes of the training and transfer to untrained contexts, and to compare UTI-VF to other methods used in L2 speech sound learning and remediation. Another limitation of the reported studies is that the UTI training sessions were delivered one-on-one, resulting in high time demands on both the learner and the trainer. The question remains how this method could be applied in a classroom setting where it is most likely that only one ultrasound machine is available. Finally, other components of L2 speech production, such as lingual articulatory setting, should be investigated.

These questions represent the core of the project "Ultrasound tongue imaging for speech sound learning and remediation" which has been running at the Institute of Phonetics, Faculty of Arts, Charles University, since September 2018. The project aims to investigate whether UTI-VF helps L2 speakers to improve the production of L2 sounds, to explore the method used in individual practice and classroom setting, to compare different methods of learning new sounds in children and adults, and to investigate the effect of articulatory training on L2 speech sound perception.

So far, we published one experiment, in which we evaluated the remediation of already acquired English /e — æ/ contrast in native Czech speakers by comparing Czech speaking English with either a Czech or English lingual articulatory setting (Kocjančič Antolík & Volín, 2019). The articulatory setting describes the positions of articulators which allow the most effortless execution of any oncoming speech sound (for review see Jenner, 2001). Because languages differ in their speech sound systems, they also differ in their articulatory setting. The studies so far have shown that lingual articulatory setting can be reliably measured in inter-speech pauses and is language-specific (Gick, Wilson, Kock & Cook, 2004), that bilinguals use two different articulatory settings (Wilson & Gick, 2014), and that L2 learners can use their native setting or exhibit some characteristics of the L2 (Święciński, 2013; Benítez, Ramanarayanan, Glodstein & Narayanan, 2014). Our study demonstrated that L2 speakers can learn the L2 lingual articulatory setting via UTI-VF and start using it comfortably and reliably in an experimental condition. Moreover, the participants who spoke with the English lingual articulatory setting reported that they sounded more English.

In our second experiment, developed together with Tomáš Bořil (Institute of Phonetics, Charles University, Prague) and Susanna Hofmann (Department of Germanic Studies, Charles University, Prague), we evaluated how UTI-VF can be applied to a typical L2 classroom setting and we compared UTI-VF to a real-time VF based on acoustic analysis of vowels (publication in preparation). Ten students of the Swedish language used the two methods to practice the production of Swedish vowels during three classroom sessions. Each student received about seven minutes of individual practice and observed the practice of their classmates. The results showed that all students improved the production but no differences between the two methods were

found. More importantly, the experiment demonstrated that VF methods can be successfully used in a classroom and that the students seem to benefit from observing not only their own productions but also those of their fellow classmates.

Our ongoing experiments further explore issues relating to UTI-VF in L2 learning, intending to propose the best way to facilitate L2 speech sound learning and remediation.

ACKNOWLEDGEMENTS

This work was supported by the OP VVV project no. CZ.02.2.69/0.0/0.0/17_050/0008466.

REFERENCES

Adler-Bock, M., Bernhardt, B. M., Gick, B., & Bacsfalvi, P. (2007). The use of ultrasound in remediation of North American English /r/ in 2 adolescents. *American Journal of Speech-Language Pathology*, 6, 128–139.

Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In A. S. Rauber, M. A. Watkins & B. O. Baptista (Eds.), *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech* (pp. 10–27). Florianópolis: Federal University of Santa Catarina.

Arteaga, D. L. (2000). Articulatory phonetics in the first-year Spanish classroom. *The Modern Language Journal*, *84*(3), 339–354.

Ballard, K. J., Smith, H. D., Paramatmuni, D., McCabe, P., Theodoros, D. G., & Murdoch, B. E. (2012). Amount of kinematic feedback affects learning of speech motor skills. *Motor Control*, *16*(1), 106–119.

Barriuso, T. A., & Hayes-Harb, R. (2018). High variability phonetic training as a bridge from research to practice. *CATESOL Journal*, *30*(1), 177–194.

Benítez, A., Ramanarayanan, V., Glodstein, L., & Narayanan, S. (2014). A real-time MRI study of articulatory setting in second language speech. *Proceedings of the Annual Conference on International Speech Communication Association INTERSPEECH*, 701–705.

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.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (pp. 13–34). Amsterdam, The Netherlands: John Benjamins Publishing.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Attention, Perception, & Psychophysics*, *61*(5), 977–985.

Cleland, J., Scobbie, J. M., & Wrench, A. A. (2015). Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics & Phonetics*, *29*(8–10), 575–597.

Cleland, J., Scobbie, J. M., Nakai, S., & Wrench, A. A. (2015). Helping children learn non-native articulations: The implications for ultrasound-based clinical intervention. In The Scottish Consortium for ICPhS 2015 (Eds.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: University of Glasgow. Paper number 698.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching:

A research-based approach. *TESOL quarterly*, *39*(3), 379–397.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1997). Pronunciation instruction for "fossilized" learners: Can it help? *Applied Language Learning*, *8*, 217–235.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, *48*, 393–410.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, (pp. 233–277). Baltimore: York Press.

Gick, B., Wilson, I., Kock, K., & Cook, C. (2004). Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica*, *61*, 220–233.

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*(2), 866–877.

Jenner, B. (2001). Articulatory setting: Geneaologies of an idea. *Historiographia Linguistica*, *28*, 121–141.

Kocjančič Antolík, T., & Volín, J. (2019). Ultrasound tongue imaging for vowel remediation in Czech English. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 3651–3655). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Kocjančič Antolík, T., Pillot-Loiseau, C., & Kamiyama, T. (2019). The effectiveness of real-time ultrasound visual feedback on tongue movements in L2 pronunciation training: Japanese learners' progress on the French vowel contrast /y/-/u/. *Journal of Second Language Pronunciation*, *5*(1), 72–97.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, *26*(2), 227–247.

Lin, S., Cychosz, M., Shen, A., & Cibelli, E. (2019). The effects of phonetic training and visual feedback on novel contrast production. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 899–903). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

McLeod, S. (2011). Speech–language pathologists' knowledge of tongue/palate contact for consonants. *Clinical Linguistics & Phonetics*, *25*(11–12), 1004–1013.

Ouni, S. (2011). Tongue gestures awareness and pronunciation training. *12th Annual Conference of the International Speech Communication Association-Interspeech 2011*, 881–884.

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472.

Preston, J. L., McCabe, P., Rivera-Campos, A., Whittle, J. L., Landry, E., & Maas, E. (2014). Ultrasound visual feedback treatment and practice variability for residual speech sound errors. *Journal of Speech, Language, and Hearing Research*, *57*(6), 2102–2115.

Roon, K. D., Kang, J., & Whalen, D. H. (2020). Effects of ultrasound familiarization on production and perception of nonnative contrasts. *Phonetica*, 1–44.

Sisinni, B., d'Apolito, S., Fivela, B. G., & Grimaldi, M. (2016). Ultrasound articulatory training for teaching pronunciation of L2 vowels. *ICT for language learning*, 265–270.

Święciński, R. (2013). An EMA study of articulatory settings in Polish speakers of English. In E. Waniek-Klomaczak & L. R. Shockey (Eds). *Teaching and Researching English Accents in Native and Non-native Speakers* (pp. 73–82). Berlin, Heidelberg: Springer-Verlag.

Tateishi, M., & Winters, S. (2013). Does ultrasound training lead to improved

perception of a non-native sound contrast? Evidence from Japanese learners of English. *Proc. 2013 annual conference of the Canadian Linguistic Association*, 1–15.

Thomson, R. I. (2018). High variability [pronunciation] training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, *4*(2), 208–231.

Thomson, R. I., & Derwing, T. M. (2016). Is phonemic training using nonsense or real words more effective? In J. Levis, H. Le, I. Lucic, E. Simpson & S. Vo (Eds.), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference* (pp. 88–97). Ames: Iowa State University.

Wilson, I., & Gick, B. (2014). Bilinguals use language-specific articulatory settings. *Journal of Speech, Language, and Hearing Research*, *23*, 361–373.

Wilson, I., Gick, B., O'Brien, M. G., Shea, C., & Archibald, J. (2006). Ultrasound technology and second language acquisition research. In *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)* (pp. 148–152). Somerville, MA: Cascadilla Proceedings Project.

Wong, J. (2012). Training the perception and production of English /e/ and /æ/ of Cantonese ESL learners: A comparison of low vs. high variability phonetic training. In *Proceedings of the 14th Australasian International Conference on Speech Science and Technology* (pp. 3–6). Canberra, Australia: Australasian Speech Science and Technology Association (ASSTA).

Tanja Kocjančič Antolík | Fonetický ústav, Filozofická fakulta, Univerzita Karlova
<tanja.kocjancicantolik@ff.cuni.cz>