Ultrasound Technology and its Role in Cantonese Pronunciation Teaching and Learning

Abstract

Over a decade of research has shown how using ultrasound imaging for biovisual feedback of tongue movement can help improve language learners' pronunciation. However, ultrasound can be challenging to implement in a classroom context, as it is typically best with small groups and requires specialized training. One possible solution is the use of ultrasound overlay videos, in which mid-sagittal ultrasound videos of tongue movements in speech are overlaid on videos of an external view of a speaker's head to create videos in which the movements of the face and the tongue are viewed simultaneously. In this paper, we report on a study investigating the use of ultrasound overlay videos as a pronunciation learning tool in Cantonese language classes. Using a blended learning paradigm, half of the students interacted with the videos online to learn about two challenging sets of sounds: unreleased final stops and low central vowels, while half interacted with audio-only media under otherwise identical conditions. Results show that students who received the ultrasound-based training performed better in perception and production tasks.

Authors

Heather Bliss
Department of Linguistics, University of Victoria
Clearihue Building D353
Victoria, BC V8W4P6
hbliss@uvic.ca

Heather Bliss is a Banting Postdoctoral Fellow at the University of Victoria. Having worked with Indigenous communities in Western Canada over the past fifteen years on language documentation projects, she has a background and interest in endangered language revitalization and community-based strategies for heritage language learning. Recently, she has been exploring the use of ultrasound technology as a tool for pronunciation teaching and learning, as well as language documentation, in Indigenous communities.

Lauretta Cheng
Department of Linguistics, University of British Columbia
Totem Field Studios, 2613 West Mall
Vancouver, BC V6T1Z4
lauretta.cheng@alumni.ubc.ca

Lauretta Cheng is an undergraduate research assistant in the Department of Linguistics at the University of British Columbia. Her research interests are on phonetics in bilingual and heritage speakers, second language teaching and acquisition, as well as Cantonese linguistics. She has worked on projects involving L2 pronunciation teaching, bilingual articulatory compensation, and phonetic variation and change in heritage Cantonese

Murray Schellenberg
Department of Linguistics, University of British Columbia
Totem Field Studios, 2613 West Mall
Vancouver, BC V6T1Z4
mhschellenberg@gmail.com

Murray Schellenberg is a post-doctoral researcher and manager of the Interdisciplinary Speech Research Laboratory (ISRL) at the University of British Columbia. His main research interests involve the interaction of language and music. He is particularly interested in issues related to the influence of language on the creation and/or performance of music. Topics of study include the realization of lexical tone in singing in tone languages (particularly in Cantonese and Mandarin); the carry-over of spoken accent to singing performance and influences of phonological structure on text setting.

Zoe Lam
Department of Linguistics, University of British Columbia
Totem Field Studios, 2613 West Mall
Vancouver, BC V6T1Z4
zoelam@alumni.ubc.ca

Zoe Lam is a PhD Candidate in Linguistics from the University of British Columbia. Her research interests include the perception of lexical tones, heritage linguistics, phonetics-phonology interface, and the grammar of Cantonese discourse particles. Currently her dissertation project focuses on the perception of Cantonese speech sounds by English-dominant heritage speakers in Canada who grew up in a Cantonese household. Zoe is also a social activist for linguistic diversity. She is an active member of the Chinatown community in Vancouver and is also the instructor of a Cantonese Saturday School.

Raymond Pai
Department of Asian Studies, University of British Columbia
607 – 1871 West Mall
Vancouver, BC V6T1Z2
kitpai@mail.ubc.ca

Raymond Pai is lecturer and the director of the Cantonese program at the University of British Columbia. He is recently named Lecturer of the Year in the Department of Asian Studies at UBC. He taught Chinese at the Defense Language Institute Foreign Language Center and the Middlebury Institute of International Studies in Monterey, California for nearly a decade. He is also the Cantonese examiner for the University of Arizona. His research focus is the language policies in Hong Kong and Cantonese language pedagogy. His other academic and research interests include language learning motivation, language attitudes, language technology and testing, and psycholinguistics.

Bryan Gick
Department of Linguistics, University of British Columbia
Totem Field Studios, 2613 West Mall
Vancouver, BC V6T1Z4
gick@mail.ubc.ca

Bryan Gick is Professor and Head of the Department of Linguistics at University of British Columbia and a Senior Researcher at Haskins Laboratories. His research interests have focused mainly on understanding the physical mechanisms of speech production and their interactions with perception, phonology and phonetics in normal, disordered, and children's speech. He has developed techniques for applying ultrasound imaging technology to speech research, protocols for studying tactile perception of speech, methods for observing low-level planning in speech motor behaviour, and computational tools for biomechanical simulation.

Introduction

There is a growing body of evidence to support the use of ultrasound imaging technology in L2 learning (Cleland et al. 2015; Gick et al. 2008; Noguchi et al. 2015; Ouni 2014; Pillot-Loiseau et al. 2015; Tateishi & Winters 2013; Tsui 2012; Wilson 2014; Wilson & Gick 2006; Wu et al. 2015). Ultrasound is useful in pronunciation teaching and learning, as it displays internal articulatory processes and facilitates the explanation and understanding of how to pronounce certain L2 sounds. However, there are various challenges associated with incorporating ultrasound into L2 learning programs. Specifically, given the current technology, ultrasound lends itself best to small groups or even individual learners. Moreover, raw, unedited ultrasound images may be difficult for untrained learners (and/or instructors) to interpret.

To address these challenges, a team of researchers at the University of British Columbia developed a technique for creating ultrasound overlay videos, in which mid-sagittal ultrasound videos of tongue movements in speech were overlaid on videos showing an external view of a speaker's head to create videos in which the movements of the face and the tongue are viewed simultaneously. The videos can be viewed at http://enunciate.arts.ubc.ca. There are 91 videos in the video library, corresponding to most of the sounds in the International Phonetic Alphabet.

Designed for application in blended or online learning paradigms, the videos have been incorporated into introductory linguistics courses as well as some language courses at the University of British Columbia.

Although ultrasound is well-established as an effective interactive biofeedback tool in L2 pronunciation learning, there is yet little research on the effectiveness of ultrasound overlay videos. Survey data suggests that learners enjoy the videos and feel they help for learning new sounds (Tsuda et al. 2015; Yamane et al. 2015). Also, a controlled experiment comparing the impact on linguistics student learning after exposure to different types of online tutorial materials found that student performance was better with tutorials that included ultrasound overlay videos than tutorials that included text materials (Abel et al. 2016). However, this current study is the first to systematically evaluate the effectiveness of ultrasound overlay videos as an L2 pronunciation learning tool.

While the video library of individual sounds is useful as a general resource, there is also a need for customized videos, in some cases targeting specific phonological contrasts in a given language, and in other cases to present a familiar face in a particular linguistic community. In this paper, we report on a pilot project to develop and evaluate ultrasound overlay videos focused on Cantonese words and sounds.

The reasons for focusing on Cantonese is twofold. First, the Cantonese language program at the University of British Columbia is new and in need of teaching materials. Launched in 2015, the program is the first and only for-credit university program of its kind in Canada. Second, the phonetic and phonological properties of Cantonese pose specific pronunciation challenges for learners, some of which are likely to benefit from ultrasound-based instruction. Specifically, with regards to consonants, unaspirated stops /p/, /t/, /k/ are unreleased in coda position and as such, there is no release burst with these consonants to provide a perceptual cue to the place of articulation (Cheung 1986). The formant frequencies of the preceding vowel are said to provide an acoustic cue to the identity of these consonants (e.g., Ciocca et al. 1994; Khouw & Ciocca 2006), but despite these acoustic cues, the unreleased stops can be difficult to distinguish, all “tending to sound like a glottal stop to an English speaker” (Matthews & Yip 2011). With regards to vowels, Cantonese has two central low vowel phonemes [a:] and [ɐ], and given their close proximity, these two vowels are difficult for learners to distinguish.

Our research question is as follows: *Does interacting with ultrasound-enhanced videos improve beginner Cantonese learners’ ability to differentiate between challenging Cantonese sounds in their perception and production?* We hypothesize that students who interact with ultrasound overlay videos will perform better in perception and production tasks that differentiate unreleased obstruents [p̚], [t̚], [k̚] and central low vowels [ɐ] and [a:] than students who interact with audio samples alone. In what follows, we describe an experiment that tested this hypothesis, beginning with a discussion of the methodology, followed by results, discussion, and concluding remarks.

Methodology

In this study, we developed ultrasound overlay videos for minimal sets of words that isolated the two contrasts, namely unreleased stops and central low vowels. We then conducted a comparative study in which half the learners were given access to the ultrasound overlay videos, and half were given access to the corresponding audio files. Learners were subsequently tested on their ability to differentiate the sounds in perception and production.

The participants were 13 undergraduate students enrolled in CNTO 301 (Basic Cantonese I) at the University of British Columbia. This course is an elementary level Cantonese language course designed for non-heritage learners with no prior exposure to or background in Cantonese. The focus of the course is on training for basic oral skills in the language. The 13 students were randomly assigned to two groups; an experimental group (n=7) received the ultrasound overlay videos and a control group (n=6) received audio files.

The test items consisted of minimal sets contrasting the vowels and consonants, as shown in Tables 1 and 2.

Table 1. Test items: Vowels

Word	**Transcription**	**English Translation**
新	s**ɐ**n1	'new'
山	s**a:**n1	'mountain'
諗	l**ɐ**m2	'think'
攬	l**a:**m2	'hug

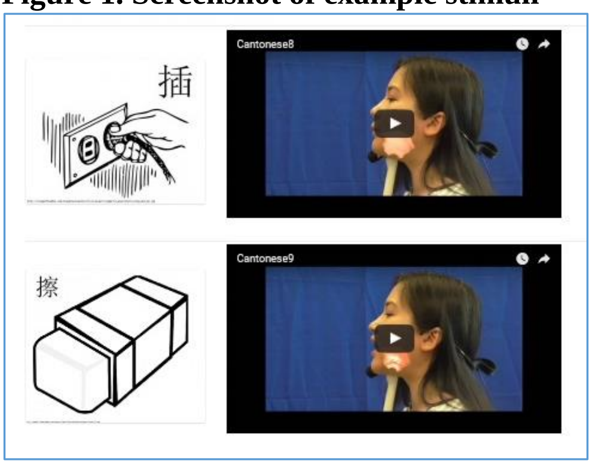
Table 2. Test items: Consonants

Word	**Transcription**	**English Translation**
濕	sɐ**p**1	'wet'
失	sɐ**t**1	'lose'
塞	sɐ**k**1	'(traffic) jam'
插	ts^ha:**p**3	'insert'
擦	ts^ha:**t**3	'erase'
拆	ts^ha:**k**3	'pull down / disassemble'

The words in Tables 1 and 2 were used to create stimuli consisting of ten videos and ten corresponding audio clips. The videos were produced by filming the side profile of the speaker (using a Panasonic Camcorder) while at the same time recording the ultrasound information from her tongue (using a CHISON portable ultrasound machine, recorded into an Mac Pro laptop using iMovie). The speaker sat with the ultrasound probe positioned under the chin and read each target word three times. Two takes were filmed, and one example of each word – deemed by a native Cantonese speaker to be the clearest production – was selected to be developed into the ultrasound overlay video. Once recorded, the videos were time-aligned and the ultrasound videos trimmed, shaded, and overlaid on the videos of the face using Adobe Premiere (see Abel et al. 2015; Yamane et al. 2015 for a detailed description of the video production methods). The audio signal from each video was extracted and saved as a wav file. In each media file (video or audio) a single token was repeated three times, and the second repetition was manipulated to be 1.5 times slower.

The learners interacted with the stimuli through two near-identical tutorial websites created using WordPress. The experimental group was given access to a site that linked to the videos, and the control group was given access to the same site, but with links to the audio clips. On both sites, the media files were displayed alongside pictures that gave a graphical depiction of each word, as well as the word itself in the Cantonese orthography. The rationale for this type of display was based on the fact that students were not trained in how to read the Cantonese orthography, and they likely would not know the meaning of the words, but the images could make it clear that the media files corresponded to real Cantonese words and represented real contrasts. An example (from the experimental group's site) is given in Figure 1.

Figure 1. Screenshot of example stimuli



Each site was comprised of five pages, including a homepage plus one page for each minimal set. The pages were accessed via hyperlinks at the top of each page labelled "Set 1," "Set 2," etc. At the top of each page, a message directed students to focus on the phonetic contrast in question, as follows: "*Watch each video to hear the word corresponding to the picture. The word will be played three times, where the second repetition is slowed down. Pay special attention to the vowel sounds in the middle of the words."* Access to the tutorial websites was through the University's Online Learning Management System.

Following a week-long period of interacting with the websites at their leisure, learners were evaluated on their perception and production of the Cantonese sounds in question. Regarding the perception evaluation, this was administered through an online quiz with ten multiple choice questions that had accompanying audio files, recorded by a second native Cantonese speaker. Half of the questions were forced choice questions, with the following phrasing: "*Listen to this word. Does it correspond to the picture on the left or on the right*?" The other half of the questions were AXB questions; two of these focused on vowels, and the remaining three on consonants. The phrasing for the latter was as follows: "*Listen to these 3 words. If the consonant at the end of the first word is the same as the one at the end of the middle word, choose 1. If the consonant at the end of the last word is the same as the one at the end of the middle word, choose 3*."

Regarding the production evaluation, both before and after the students accessed the tutorial websites, we recorded their productions of the numerals one through ten in both random and sequential orders. The recordings were carried out in a quiet classroom using a USB microphone and Audacity 2.1.2, and the recording conditions were identical in both the pre- and post-recording sessions. The numerals were chosen because the students were familiar with these words and therefore would not need to repeat after their instructor or another native speaker, and the numerals contain the relevant phonological contrasts, as illustrated in Table 3.

Table 3. Numerals illustrating relevant contrasts

Vowels	3 三 [s**aː**m1]	10 十 [s**ɐ**p6]	
Vowels	8 八 [b**aː**t3]	7 七 [tsʰ**ɐ**t1]	

Consonants	10 十 [sɐ**p**6]	7 七 [tsʰɐ**t**1]	6. 六 [lo**k**6]

With respect to data analysis, the perception data was analysed according to the percentage of correct responses by question type across conditions. For the production data, we performed both acoustic and rater analyses. Acoustic analysis was carried out using Praat (Boersma and Weenick 2015), and analysis of the vowels entailed collecting F1 values of the vowels at midpoint and comparing across vowels and conditions. Analysis of the consonants adapted a method used by Ciocca et al. 1994 and Khouw and Ciocca 2006, and entailed calculating the difference between F2 and F3 values from the midpoint to the endpoint of vowel duration, and comparing across pre- and post-recordings for each subject and across conditions. The rater analysis entailed having four native Cantonese speakers rate the nativeness of the vowels and consonants separately on scale of one to five and comparing mean rater scores for vowels and consonants.

Results

The results of the perception quiz are presented in Table 4.

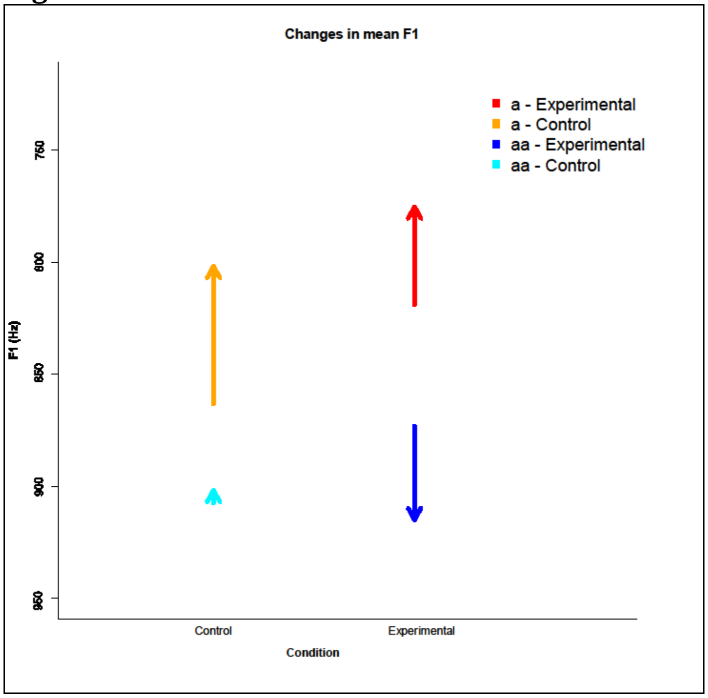
Table 4. Perception Results

Group	**Mean Correct Responses (AXB)**	**Mean Correct Responses (Forced Choice)**
Experimental	88% (n=5)	70% (n=6)
Control	77% (n=6)	63% (n=6)

Statistical analysis of the perception results is not reliable due to the small number of participants. We omitted one outlier whose score was significantly lower than all other participants, as well as a participant from the experimental group who did not answer the AXB questions.

The results of the acoustic analysis of the vowels are presented in Figure 2 (in which “aa” refers to [a:] and “a” refers to [ɐ]).

Figure 2. Acoustic measures of the vowels



The arrows in Figure 1 show the mean direction of change in F1 values for each vowel from pre- to post-recording. While both control and experimental groups seem to have a larger difference between the F1 values of the two vowels after training, the experimental group show trends towards a larger increased differentiation of the two low vowels than the control. In other words, comparing across time points, the differences between the height of the tongue for [ɐ] and [a:] is more pronounced in the post-recording sessions regardless of condition. Comparing across conditions, the experimental group appears to improve more.

The result of the rater analysis for the vowels are presented in Table 5.

Table 5. Rater Analysis (Vowels)

Group		Mean Rating (all)	Mean Rating [ɐ]	Mean Rating [a:]
Control	pre-	3.565	3.563	3.568
	post-	3.456	3.573	3.339
Experimental	pre-	3.784	3.755	3.813
	post-	3.456	3.568	4.078

Native	--	4.555	4.650	4.517

Two-way random intra-class correlations were run as a measure interrater reliability, indicating fair agreement amongst the four raters [ICC(2, 4) = 0.455]. Table 5 shows that there were significant differences between ratings of native speakers and learners, but not between conditions in the Cantonese learners. Nevertheless, there appears to be a slight trend towards an increase in nativeness rating for the experimental group's production of [a:], which matches the trend of improvement from acoustic data.

Turning to the consonants, we found no significant differences in the F2 or F3 values between the different consonants, even for model speaker. Similarly, there were no significant changes in the ratings for either group over time, as shown in Table 6. As with the vowels, measures of interrater reliability indicate that there was fair agreement amongst the four raters [ICC(2, 4) = 0.593]. While there were no trends between conditions, the data did show a pattern of nativeness ratings across the three stops whereby [p˺] was consistently rated highest and [k˺] lowest.

Table 5. Rater Analysis (Consonants)

Group		**Mean Rating (all)**	**Mean Rating [p]**	**Mean Rating [t]**	**Mean Rating [k]**
Control	pre-	3.607	4.104	3.487	3.596
	post-	3.662	3.729	3.544	3.281
Experimental	pre-	3.848	3.927	3.755	3.573
	post-	3.960	3.927	3.800	3.604
Native	--	4.450	4.750	4.156	4.594

Discussion

Recall our prediction that students who interact with the ultrasound overlay videos will perform better in production and perception of the unreleased stops and low central vowels than students who interact with audio files. While the small sample size in this study limits our ability to draw any firm conclusions, in a broad sense this prediction is borne out. The perception data trended in the right direction, with the experimental group showing more improvements than the control group, and the production data showing more mixed results, but with trends in the right direction (for the vowels particularly).

Regarding the perception data, we observed a smaller difference between the experimental and control groups with the forced choice questions, and we speculate that this may be because these questions did not test pronunciation so much as vocabulary; these relied on students' ability to recall the training words. The overall weak effect in the perception data could be due to problems with both quiz design and delivery. With respect to quiz design, instead of forced choice questions, we speculate same/different types of questions ("*Are these two sounds the same or different*?") may have been more effective. With respect to quiz delivery, a technical glitch required that students type their responses, rather than use a simpler multiple choice format, and this may have impacted the results.

As for the acoustic results, as evidenced in Figure 2, the vowels trended in the right direction, with a clearer distinction in the experimental group between the F1 values of [aː] versus [ɐ]. The consonants, on the other hand, showed no acoustic differences between [p̚], [t̚], and [k̚], even for the model speaker. One possible explanation for this lack of a distinction is an ongoing sound change in Cantonese whereby younger speakers are fluctuating between word-final [t̚] and [k̚], using the two in free variation (Law et al. 2013; To et al. 2015). It is possible that the model speaker (a younger speaker) may have not differentiated these sounds as clearly, which would have been reflected in stimuli the students received. Alternatively, it could be that glottalization obscured the contrast. Particularly with younger speakers, Cantonese final unreleased stops can often have a glottal closure component, which means that the vocal folds close to stop airflow momentarily

(Yip 2015). The [p̚], [t̚], and [k̚] stops are often produced with glottal closure, before or during tongue movement, and sometimes even instead of any tongue articulation. Such a glottal closure would at least partially obscure the acoustic cues measured. If the tongue articulation was also weaker because of the glottal closure, this could explain the weak effects in even our model speaker data, which in turn would affect the participants' performance. The consequence of this extra articulation may be that ultrasound information alone is insufficient for learning these sounds.

Regarding the results of the native speaker ratings, the lack of a clear result may be related to task effects. Raters reported that being asked to judge only the vowel or consonant when hearing the full word to be a difficult task. Consequently, it could be the case that the ratings were inadvertently influenced by overall nativeness judgments, and/or from the interference of tone. Moreover, validity and interrater reliability may have been improved by allowing raters to listen to each token multiple times.

Some other factors may have impacted the results that we found. A software glitch prevented us from tracking the amount of time each student spent on the training website, and as a result we are unaware of any correlations between time spent on the websites and production and perception performance. Furthermore, the students had taken a midterm in which they were tested on the numerals the day before the pre-recording session. The post-recording session was one week later. This timing may have influenced the students' production performance, as they would have been more well-studied at the time of the pre-recording session.

Conclusions and Future Directions

In this study, we tested the use of ultrasound overlay videos as a pronunciation learning tool, in contrast with audio-only media. While the study focused on only a small set of sounds and had a small number of participants, the results are promising insofar as they trend towards providing empirical support for the use ultrasound overlay videos for learning certain Cantonese sounds that are reportedly challenging for L2 learners. This study serves a pilot, and we are in the process of conducting a follow-up study with Cantonese learners that addresses some of the problems identified of the current study, and includes a larger number of participants. We are also developing customized videos for other widely taught languages at the University of British Columbia, including Mandarin, Spanish, French, and German, and we are responding to interest from other linguistic communities and making customized videos for languages including SENĆOŦEN, Blackfoot, Secwepemc, and Halq’emeylem (Bliss et al. 2016). As our video library grows, we continue to develop plans to integrate the videos into classrooms and other language learning environments, and to evaluate these videos’ impact on pronunciation learning.

Acknowledgements

Sincere thanks to Strang Burton, Johnathan Chan, Deena Hymowitz, Matthew Law, Yik Tung Wong for their assistance with this project, to other members of the Interdisciplinary Speech Research Laboratory and the audience at PSLLT 8 for insightful comments and feedback, and to the students of CNTO 301 (Summer 2016) for their willing participation. This work has been supported by an NIH Grant DC-02717 to Haskins Laboratories, the UBC Teaching Learning Enhancement Fund, a UBC Arts Undergraduate Research Award, and a Robert and Kazuko Barker Award.

References

Abel, J., B. Allen, S. Burton, M. Kazama, M. Noguchi, A. Tsuda, N. Yamane, & B. Gick. (2015). Ultrasound-Enhanced Multimodal Approaches to Pronunciation Teaching and Learning. *Canadian Acoustics* 43(3): 124-25.

Abel, J., H. Bliss, M. Noguchi, M. Schellenberg, N. Yamane, and B. Gick. *Instructional Reinforcements in Phonetics Pedagogy: A Comparative Study*. (2016). Unpublished manuscript (under review), the University of British Columbia.

Bliss, H., S. Burton, and B. Gick. (2016). Ultrasound overlay videos and their application in Indigenous language learning and revitalization. *Canadian Acoustics* 44(3): 136-37.

Boersma, P. & Weenink, D. (2015). Praat: doing phonetics by computer. Version 5.4.08, Available from http://praat.org/

Cheung, K.H. (1986). The phonology of present-day Cantonese. PhD dissertation, University College London.

Ciocca, V., Wong, L., & So, L. K. H. (1994). An acoustic analysis of unreleased stop consonants in word final positions. Proceedings of the International Conference on Spoken Language Processing, Yokohama, 21, 1131–1134.

Cleland, J., Scobbie, J. M., Nakai, S., & Wrench, A. (2015). Helping children learn non-native articulations: the implications for ultrasound-based clinical intervention. Paper presented at the 2015 International Conference of Phonetic Sciences, Glasgow, Scotland. Retrieved on 12 August 2015 from http://www.icphs2015.info/pdfs/Papers/ICPHS0698.pdf.

Gick, B., Bernhardt, B., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. In J. G. Hansen Edwards & M. L. Zampini (Eds.). *Phonology and second language acquisition* (pp. 309-322). Amsterdam: John Benjamins.

Khouw, E., & Ciocca, V. (2006). An acoustic and perceptual study of final stops produced by profoundly hearing impaired adolescents. *Journal of Speech, Language, and Hearing Research*, *49*(1), 172-185.

Law, S. P., Fung, R. S. Y., & Bauer, R. S. (2013). Perception and production of Cantonese consonant endings. *Asia Pacific Journal of Speech, Language and Hearing*.

Matthews, S., & Yip, V. (2011). Unbalanced bilingual acquisition as a mechanism of grammatical change. *Bilingualism: Language and Cognition, 14*(02), 159-161.

Noguchi, M., N. Yamane, A. Tsuda, M. Kazama, B. Kim, and B. Gick. (2015). Towards protocols for L2 pronunciation training using ultrasound imaging. Poster presented at PSLLT 2015. Dallas, Texas.

Ouni, S. (2014). Tongue control and its implication in pronunciation training. *Computer Assisted Language Learning*, *27*(5), 439-453.

Pillot-Loiseau, C., Kamiyama, T., & Kocjančič Antolík, T. (2015). French /y/-/u/ contrast in Japanese learners with/without ultrasound feedback: vowels, non-words and words. Paper presented at the 2015 International Conference of Phonetic Sciences, Glasgow, Scotland. Retrieved on 12 August 2015 from http://www.icphs2015.info/pdfs/Papers/ICPHS0485.pdf.

Tateishi, M., & Winters, S. (2013). Does ultrasound training lead to improved perception of a non-native sound contrast? Evidence from Japanese learners of English. Paper presented at the 2013 meeting of the Canadian Linguistic Association, Victoria, BC, Canada. Retrieved on 12 August 2015 from http://homes.chass.utoronto.ca/~cla-acl/actes2013/Tateishi_and_Winters-2013.pdf.

To, C. K. S., Mcleod, S. & Cheung, P. S. P. (2015) Phonetic variations and sound changes in Hong Kong Cantonese: Diachronic review, synchronic study and implications for speech sound assessment. *Clinical Linguistics & Phonetic*s, 29(5), 333-353.

Tsuda, A., Kim, B., Gick, B, Kazama, M., Yamane, N, & Burton, S. (2015). *Ultrasound-integrated pronunciation tutorials*. Roundtable discussion presented at the Society for Teaching and Learning in Higher Education, Vancouver, BC.

Tsui, H. M. (2012). *Ultrasound speech training for Japanese adults learning English as a second language*. Unpublished MSc thesis, University of British Columbia.

Wilson, I. (2014). Using ultrasound for teaching and researching articulation. *Acoustical Science and Technology*, *35*(6), 285-289.

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

Wu, Y., Gendrot, C., Hallé, P., & Adda-Decker, M. (2015). On improving the pronunciation of French /r/ in Chinese learners by using real-time ultrasound visualization. Paper presented at the 2015 International Conference of Phonetic Sciences, Glasgow, Scotland. Retrieved on 12 August 2015 from http://www.icphs2015.info/pdfs/Papers/ICPHS0786.pdf.

Yamane, N., J. Abel, B. Allen, S. Burton, M. Kazama, M. Noguchi, A. Tsuda, and B. Gick. (2015). Ultrasound-integrated pronunciation teaching and learning. Paper presented at Ultrafest VII, University of Hong Kong.

Yip, J. (2015). Gestural reduction of Hong Kong Cantonese syllable-final oral stops. Proceedings of the International Conference of Phonetic Sciences, Glasgow, Scotland. Retrieved on 11 August 2016 from http://www.icphs2015.info/pdfs/Papers/ICPHS0912.pdf.