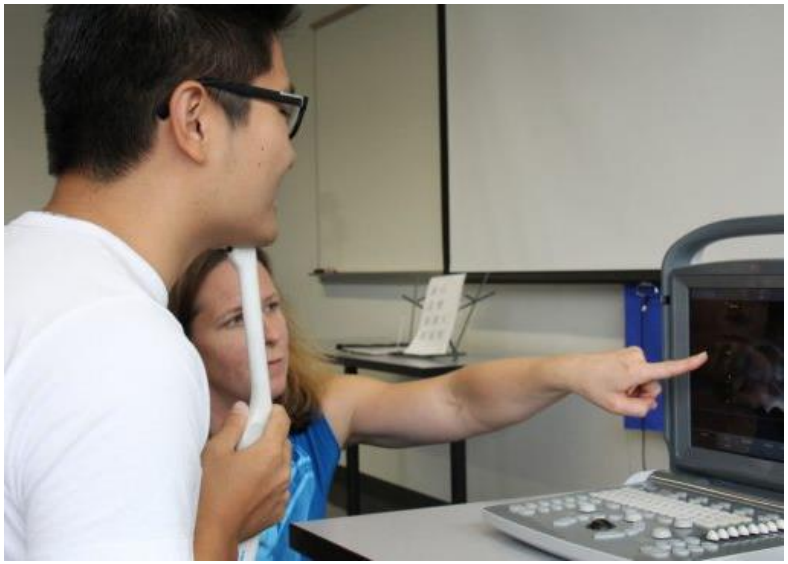
# Ultrasound Technology and its Role in Cantonese Pronunciation Teaching and Learning

Heather Bliss, Lauretta Cheng,
Murray Schellenberg, Zoe Lam,
Raymond Pai, & Bryan Gick
*University of British Columbia*



# **Background:** Ultrasound in L2 pronunciation instruction

- 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 displays internal articulatory processes and facilitates the explanation and understanding of how to pronounce challenging sounds

# Challenges of using ultrasound in L2 teaching

- **<u>The challenges:</u>**
  - Ultrasound lends itself best to small groups or even individual learners
  - Unedited ultrasound images may be difficult for untrained learners (and/or instructors) to interpret

- **<u>The solution: *eNunciate*</u>**
  - Ultrasound overlay videos
  - Mid-sagittal ultrasound videos of tongue movements in speech overlaid on videos showing an external view of a speaker's head
  - Movements of the face and the tongue are viewed simultaneously.
  - Videos can be used in blended or independent learning paradigms

# Effectiveness of ultrasound overlay

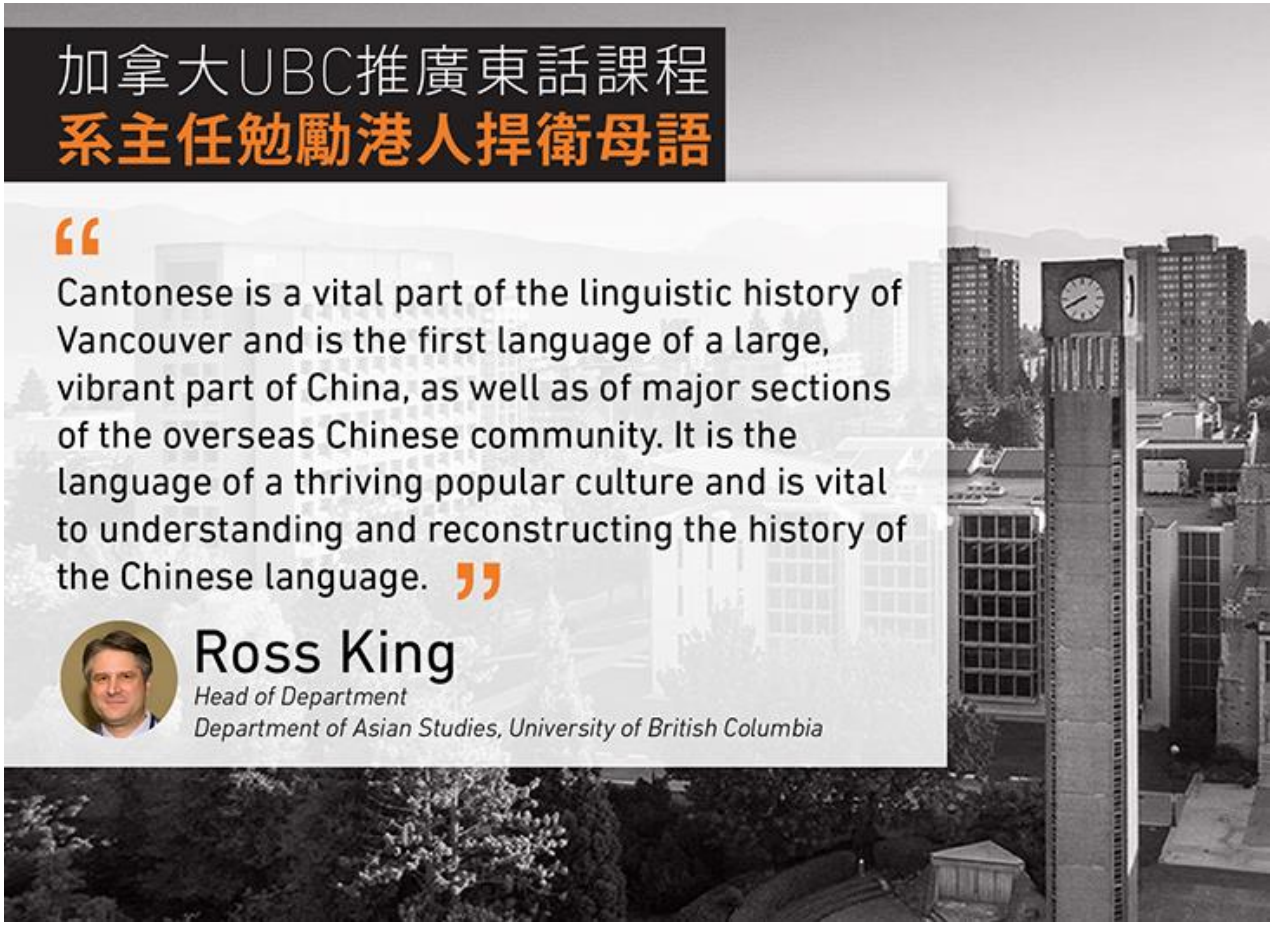
- Although ultrasound is well-established as an effective interactive biofeedback tool in L2 pronunciation learning …
- … There is little research on the effectiveness of ultrasound overlay videos
  - Survey data suggests that learners enjoy the videos and feel they help for learning new sounds (Tsuda et al. 2015; Yamane et al. 2015).
  - One controlled experiment with introductory linguistics students found that student performance was better with online tutorials that included ultrasound overlay videos than tutorials that included text materials (Abel et al., in prep).

- This current study is the first to systematically evaluate the effectiveness of ultrasound overlay videos as an L2 pronunciation learning tool

# Customizing ultrasound overlay videos

- 91 ultrasound overlay videos in our library (for each sound in the IPA)
- Useful as a general resource, but also a need for customized videos
  - To target specific phonological contrasts
  - To present a familiar face in a particular linguistic community
- Pilot project: Cantonese ultrasound overlay videos

# Cantonese




- Launched at UBC in 2015, the Cantonese language program is the first and only for-credit university program of its kind in Canada
- Phonetic and phonological properties of Cantonese pose specific pronunciation challenges for learners
- New program: learning tools needed

# Challenging Cantonese sounds

**Consonants:**

- Unaspirated stops /p/, /t/, /k/ are unreleased in coda position
- No release burst to provide a perceptual cue to the place of articulation (Cheung 1986)
- Formant frequencies of the preceding vowel are said to provide an acoustic cue (e.g Ciocca et al. 1994; Khouw & Ciocca 2006).
- Despite acoustic cues in the vowels, the unreleased stops can be difficult to distinguish, all "tending to sound like a glottal stop to an English speaker" (Matthews & Yip 2011)

**Vowels:**

- Two central low vowel phonemes [aː] and [ɐ]
- Given their close proximity, these two vowels are difficult for learners to distinguish

# Research Question

- *Does interacting with ultrasound-enhanced videos improve beginner Cantonese learners' ability to differentiate between challenging Cantonese sounds in their perception and production?*

**Hypothesis:**

- Students who interact with ultrasound overlay videos will perform better in perception and production tasks that differentiate unreleased obstruents [pˀ], [tˀ], [kˀ] and central low vowels [ɐ] and [aː] than students who interact with audio samples alone

# Methodology

# Methods

- We developed ultrasound overlay videos for minimal sets of words that isolated the two contrasts (unreleased stops, central low vowels)
- We 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 tested on their ability to differentiate the sounds in perception and production

# Participants

- 13 undergraduate students
- Enrolled in CNTO 301 (Basic Cantonese I)
  - Elementary level (part I) Cantonese language course
  - Non-heritage learners with no prior exposure to or background in Cantonese
  - Focus on training for basic oral skills in Cantonese
- Randomly assigned to two groups:
  - Experimental group (n = 7) received ultrasound overlay videos
  - Control group (n = 6) received audio files

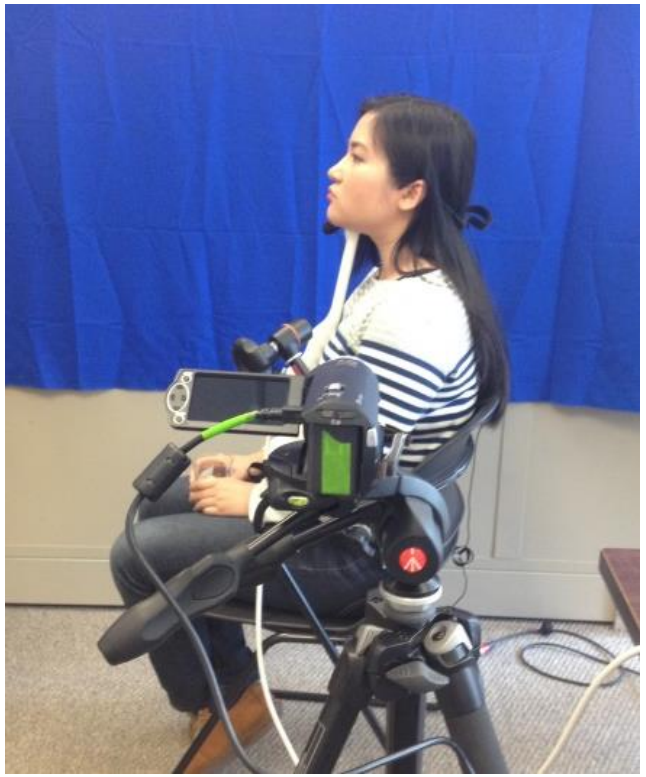
# Materials: Test items

- Minimal sets contrasting the vowels and the obstruents (two for each contrast)

| | Transcription | English gloss |
|---|---|---|
| 新 | sɐn1 | ‘new’ |
| 山 | san1 | ‘mountain’ |
| 諗 | lɐm2 | ‘think’ |
| 攬 | lam2 | ‘hug’ |

| | Transcription | English gloss |
|---|---|---|
| 濕 | sɐp1 | ‘wet’ |
| 失 | sɐt1 | ‘lose’ |
| 塞 | sɐk1 | ‘(traffic) jam’ |
| 插 | tsʰap3 | ‘insert’ |
| 擦 | tsʰat3 | ‘erase’ |
| 拆 | tsʰak3 | ‘pull down/disassemble |

# Materials: Stimuli



- 10 ultrasound overlay videos

- **Production procedure:**
- Native Cantonese speaker read the words (in the orthography) six times each. Best tokens selected.

- Videos time-aligned; ultrasound videos trimmed, shaded, overlaid on video of the face using Adobe Premiere (procedure described in detail by Abel et al. 2015; Yamane et al. 2015).

- Audio files extracted and saved as wav files.

# Materials: Tutorial websites

- Two near-identical websites were created using WordPress
  - www.blogs.ubc.ca/cantoneseproduction → linked to video
  - www.blogs.ubc.ca/productioncantonese → linked to audio

- Each media file was displayed alongside a picture that gave a graphical depiction of the word, as well as the word itself in the Cantonese orthography

- Each site was comprised of 5 pages:
  - Homepage plus one page for each minimal set
  - Pages accessed via hyperlinks at top of page labelled Set 1, Set 2, etc.




# Materials: Perception quiz


- 10 multiple choice questions (with accompanying audio files, recorded by a different native Cantonese speaker)

- 5 forced choice questions
  - “Listen to this word. Does it correspond to the picture on the left or on the right?”

- 5 AXB questions (2 for vowels; 3 for consonants)
  - “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. ”

# Procedure (Experiment)

- Participants were given access to the tutorial websites through UBC's Learning Management System ("Connect")
  - They were given one week to interact with the tutorial and complete the online quiz (also administered through Connect)

- Participants' productions recorded before and after accessing tutorials
  - Numerals 1-10 recorded in random and sequential orders
  - Students were familiar with these words
  - The numerals contain the relevant contrasts
    - Vowels: 3 三 [s**a**m1] vs. 10 十 [s**ɐ**p6]
    - Vowels: 8 八 [b**a**t3] vs. 7 七 [ts$^{h}$**ɐ**t1]
    - Consonants: 10 十 [sɐ**p**6] vs. 7 七 [ts$^{h}$ɐ**t**1] vs 6. 六 [lo**k**6]

# Procedure (Analysis)

- **<u>Perception data:</u>**
  - Percentage correct by question type across conditions
- **<u>Production data:</u>**
  - ***Acoustic analysis***
    - Vowels: collected formant values from midpoint
      - Compared across vowels ([ɐ] and [a:]) and condition
    - Consonants: calculated difference between F2 and F3 values from midpoint to endpoint of vowel duration
      - Compared across pre- and post-recordings for each subject, then across conditions
  - ***Rater analysis***
    - 4 native Cantonese speakers asked to rate nativeness of vowels and consonants separately on scale of 1 to 5
      - Compared mean rater scores for vowels and consonants

# Results

# Results – Perception (quiz)

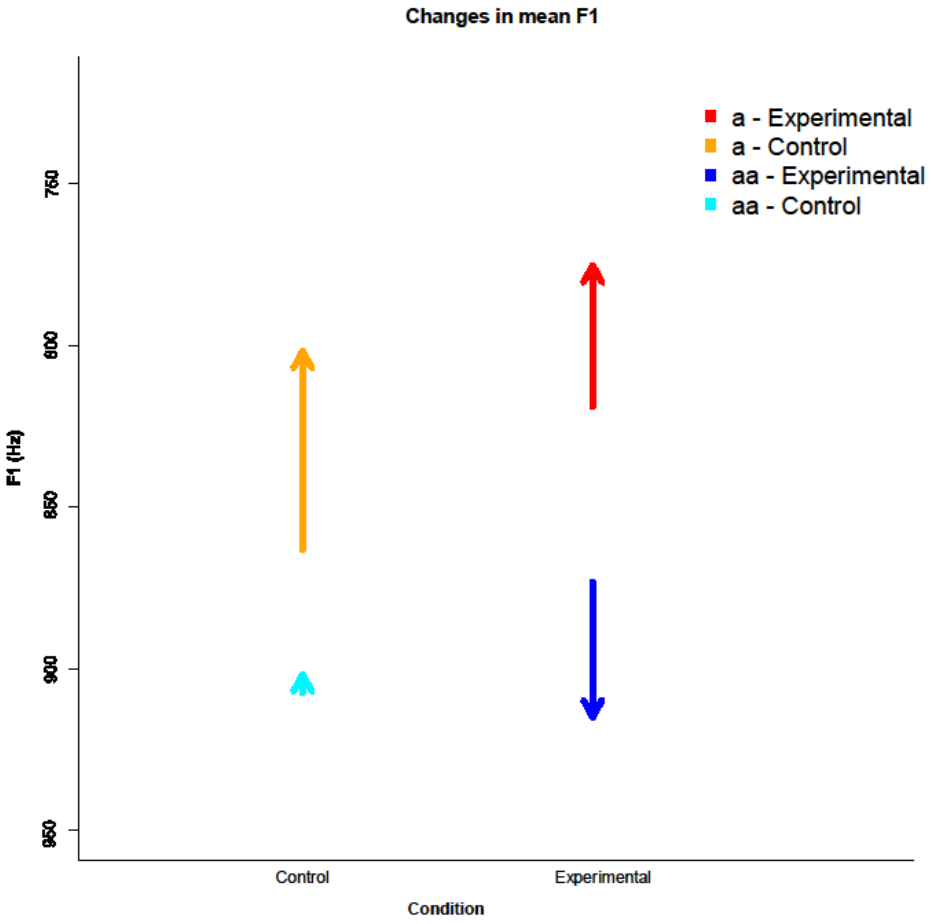
| Group | Mean Correct Responses (AXB) | Mean Correct Responses (Forced Choice) |
|---|---|---|
| Experimental | **88%** (n = 5) | **70%** (n = 6) |
| Control | **77%** (n = 6) | **63%** (n = 6) |

*omitted one outlier whose score was significantly lower than all other participants
*omitted one participant from experimental group who did not answer AXB questions

(statistical analysis not reliable due to small number of participants)

# Results – Production (vowels)

**Acoustic measures**: Experimental but not control group showed trends towards increased differentiation of the vowels in F1 values



# Results – Production (vowels), continued

- **Interrater reliability**: fair agreement [ICC(2, 4) = 0.455]
- **Rater data:** Not much change based on ratings
  - Slight trend towards increase in rating for Experimental group [a:] which matches trend from acoustic data

| 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 |

# Results – Production (consonants)

- **Acoustic measures**: No significant differences found in F2/F3 values between the different consonants – even for model speaker

# Results – Production (consonants), continued

- **Interrater reliability**: fair agreement [ICC(2, 4) = 0.593]
- **Rater data:** Not much change based on ratings
  - overall, [p] rated higher than [t]; and [t ] rated higher than [k]

| 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 | 3.80 | 3.604 |
| Native | -- | 4.450 | 4.75 | 4.156 | 4.594 |

# Discussion

# Predictions revisited

- (Small sample size limits our ability to draw firm conclusions)

- Perception data trend in the right direction (experimental group showed more improvements than control group)
- Production data show more mixed results, but trends for vowels in the right direction

# Perception data commentary

- Smaller difference with forced choice questions
  - These may not have tested pronunciation so much as memory of training words
- Weak effect may be due to:
  - Quiz design
    - Instead of forced choice, could use different types of questions, e.g. “Are these sounds the same or different?”
  - Quiz delivery
    - answers had to be typed, rather than selected in multiple choice format

# Acoustic data commentary

- Vowels trend in the right direction (but small sample size)
  - [aː] was produced higher (lower F1) while [ɐ] was lower (higher F1) - clearer distinction in experimental group

- No acoustic differences between the 3 consonants – even for the model speaker
  - Sound change in progress (incomplete) of [t] → [k] and [k] → [t] (To, McLeod & Cheung, 2015)
    - Younger Cantonese speakers found to interchange [t] and [k] endings - alveolarization and glottalization – so no F2 contrasts (Law, Fung & Bauer, 2001)
      - May have led to less clear differentiation in ultrasound overlay videos if present in model speaker

# Utility of ultrasound for Cantonese final stops

- Cantonese final unreleased stops can have a glottal component (Yip 2015)

- Especially for younger speakers, stops are often produced with glottal closure, before or during tongue movement
  - Sometimes oral stop completely replaced by glottal stop
- May make it difficult for learners to perceive differences
  - glottal stop would obscure acoustic cues
  - stop articulation may be reduced in speakers as well

  - Consequence: Ultrasound information alone may be insufficient for learning to produce these sounds

# Rater data commentary

- Rating task issues:
  - Raters found that being asked to judge only the vowel or final consonant when hearing the full word to be a difficult task
    - Ratings may have been inadvertently influenced by overall nativeness judgments (especially tone)
      - Validity and interrater reliability may be improved by allowing multiple listens

# Additional factors

- Unknown variable: time spent on website
  - We attempted to track this, but a technical glitch prevented us from gathering this data

- Midterm exam
  - Students were tested on their oral production of the numerals 1-10 in the class before the pre-recording session, but not again later.
    - This may have led to improved performance in the pre- but not post-recording

# Conclusions and further research

- Overall, the study provides empirical support for the use of ultrasound overlay videos as a L2 pronunciation learning tool
- However, the results are not as robust as we might like
- We plan to run the study again in the fall semester when course enrolment is higher and we can therefore have more participants
- We are also making customized videos for other languages (SENĆOŦEN, Blackfoot, Secwepemc, Halq'emeylem) and have plans to evaluate impact on pronunciation learning (see Bliss et al., to appear)

# 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 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 -2717 to Haskins Laboratories, the UBC Teaching Learning Enhancement Fund, a UBC Arts Undergraduate Research Award, and a Robert and Kazuko Barker Award

# References


Abel, Jennifer, Blake Allen, Strang Burton, Misuzu Kazama, Masaki Noguchi, Asami Tsuda, Noriko Yamane, & Bryan Gick. (2015). Ultrasound-Enhanced Multimodal Approaches to Pronunciation Teaching and Learning. *Canadian Acoustics* 43.

Bliss, Heather, Strang Burton, and Bryan Gick. to appear. Ultrasound overlay videos and their application in Indigenous language learning and revitalization. To appear in *Canadian Acoustics*.

Cheung, Kwan Hin. 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, Masaki, Noriko Yamane, Asami Tsuda, Misuzu Kazama, Bosung Kim, and Bryan 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.

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, Noriko, Jennifer Abel, Blake Allen, Strang Burton, Misuzu Kazama, Masaki Noguchi, Asami Tsuda, and Bryan Gick. 2015. *Ultrasound-integrated pronunciation teaching and learning*. Ultrafest VII, University of Hong Kong.

Yip, Jonathan. (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.

# Contact information

**Slides will be posted to the eNunciate project website (http://enunciate.arts.ubc.ca)**

**Questions and comments can be directed to hbliss@uvic.ca**