Ultrafest VII

The University of Hong Kong

December 8-10, 2015

ABSTRACTS

PROGRAMME OUTLINE

TUESDAY, DECEMBER 8	
9:00-9:45	**REGISTRATION & BREAKFAST**
9:45-10:00	**WECOME ADDRESS**
10:00-11:00	**SESSION ONE**
11:00-11:30	**COFFEE & TEA**
11:30-12:30	**SESSION TWO**
12:30-1:30	**LUNCH**
1:30-2:30	**KEYNOTE 1**
2:30-3:00	**DISCUSSION**
3:00-3:30	**COFFEE & TEA**
3:30-5:00	**SESSION THREE**
5:00-5:30	**GENERAL DISCUSSION I**

WEDNESDAY, DECEMBER 9	
8:30-9:00	**BREAKFAST**
9:00-10:30	**SESSION FOUR**
10:30-11:00	**COFFEE & TEA**
11:00-12:00	**SESSION FIVE**
12:00-12:30	**GENERAL DISCUSSION II**
12:30-1:30	**LUNCH**
1:30-2:30	**KEYNOTE 2**
2:30-3:00	**DISCUSSION**
3:00-3:15	**COFFEE & TEA**
3:15-5:15	**POSTER SESSION**
5:15-6:00	**BREAK**
6:00-8:00	**DINNER RECEPTION (University Lodge)**

THURSDAY, DECEMBER 10	
8:30-9:00	**BREAKFAST**
9:00-10:30	**SESSION SIX**
10:30-11:00	**COFFEE & TEA**
11:00-12:30	**SESSION SEVEN**
12:30-1:00	**GENERAL DISCUSSION III**
1:00-1:30	**BREAK**
1:30-2:30	**DIM SUM LUNCH (location: Victoria Harbour Restaurant)**

7:30-9:30	**OPTIONAL DINNER OUTING: A Symphony of Lights (Harbour Cruise - Bauhinia at the North Point Ferry Pier)**

Contents

PROGRAMME OUTLINE . . . . . **i**

PROGRAMME . . . . . **vi**

ORAL PRESENTATIONS

Tuesday December 8, 9:45-11 . . . . . **2**

Applying a 3D biomechanical model to 2D ultrasound data **2**
Alan Wrench & Peter Balch

Effect of a fixed ultrasound probe on jaw movement during speech **4**
Julián Villegas, Ian Wilson, Yuki Iguro, & Donna Erickson

Tuesday December 8, 11:30-12

Sonographic & Optical Linguo-Labial Articulation Recording system (SOLLAR) **8**
Aude Noiray, Jan Ries, & Mark Tiede

Extraction of Persian coronal stops from ultrasound images using linear discriminant analysis **10**
Reza Falahati, & Vahid Abolghasemi

Tuesday December 8, 1:30-2:30

Acoustic sensitivity of the vocal tract as a guide to understanding articulation **13**
Brad Story

Tuesday December 8, 3:30-5

Development of lingual articulations among Cantonese-speaking children **14**
Jonathan Yip & Diana Archangeli & Carol K.S. To

Speech stability, coarticulation, and speech errors in a large number of talkers . **16**
Stefan A. Frisch, Alissa J. Belmont, Karen Reddick, & Nathan D. Maxfield

Using ultrasound tongue imaging to study the transfer of covert articulatory information in coda /r/ **17**
Eleanor Lawson, James M. Scobbie, & Jane Stuart-Smith

Wednesday December 9, 9-10:30

Coarticulatory effects on lingual articulations in the production of Cantonese syllable-final oral stops **19**
Jonathan Yip

The role of the tongue root in phonation of American English stops **21**

Suzy Ahn

Bolstering phonological fieldwork with ultrasound: lenition and approximants in Iwaidja **23**
Robert Mailhammer, Mark Harvey, Tonya Agostini, & Jason A. Shaw

Wednesday December 9, 11-12

Timing of front and back releases in coronal click consonants **25**
Amanda Miller

Acoustic and articulatory speech reaction times with tongue iltrasound: What moves first? **27**
Pertti Palo, Sonja Schaeffler & James M. Scobbie

Wednesday December 9, 1:30-2:30

Neurophysiology of speech perception: Plasticity and stages of processing . . . **31**
Patrick Wong

Thursday December 10, 9-10:30

/r/-allophony and gemination: An ultrasound study of gestural blending in Dutch **32**
Patrycja Strycharczuk & Koen Sebregts

Allophonic variation: An articulatory perspective **34**
Alessandro Vietti, Lorenzo Spreafico, Vincenzo Galatà

Taps vs. palatalized taps in Japanese **36**
Noriko Yamane & Phil Howson

Thursday December 10, 11-12:30

Russian palatalization, tongue-shape complexity measures, and shape-based segment classification **39**
Kevin D. Roon, Katherine M. Dawson, Mark K. Tiede, & D. H. Whalen

Exploring the relationship between tongue shape complexity and coarticulatory resistance **40**
D. H. Whalen, Kevin D. Roon, Katherine M. Dawson, & Mark K. Tiede

An investigation of lingual coarticulation resistance using ultrasound **41**
Daniel Recasens & Clara Rodríguez

POSTERS

1. Tongue shape dynamics in swallowing **45**
Mai Ohkubo & James M Scobbie

2. Recordings of Australian English and Central Arrernte using the EchoBlaster and AAA **47**
Marija Tabain & Richard Beare

3. The effects of blindness on the development of articulatory movements in children **48**
Pamela Trudeau-Fisette, Christine Turgeon, Marie Bellavance-Courtemanche, & Lucie Ménard

4. An EPG + UTI study of syllable onset and coda coordination and coarticulation in Italian **49**
Cheng Chen, Chiara Celata, Irene Ricci, Chiara Bertini and Reza Falahati

5. A Kinect 2.0 system to track and correct head-to-probe misalignment **52**
Samuel Johnston, Rolando Coto, & Diana Archangeli

6. Articulatory settings of Japanese-English bilinguals **54**
Ian Wilson, Yuki Iguro, & Julián Villegas

7. The UltraPhonix project: Ultrasound visual biofeedback for heterogeneous persistent speech sound disorders **55**
Joanne Cleland, James M. Scobbie, Zoe Roxburgh, & Cornelia Heyde

8. Gradient acquisition of velars via ultrasound visual biofeedback therapy for persistent velar fronting **57**
Joanne Cleland, James M. Scobbie, Jenny Isles, & Kathleen Alexander

9. A non-parametric approach to functional ultrasound data: A preliminary evaluation **59**
Alessandro Vietti, Alessia Pini, Simone Vantini, Lorenzo Spreafico, Vincenzo Galatà

10. Effects of phrasal accent on tongue movement in Slovak **62**
Lia Saki Bučar Shigemori Marianne Pouplier Štefan Beňuš

11. GetContours: an interactive tongue surface extraction tool **65**
Mark Tiede & D.H. Whalen

12. The dark side of the tongue: the feasibility of ultrasound imaging in the acquisition of English dark /l/ in French learners **67**
Hannah King & Emmanuel Ferragne

13. Searching for closure: Seeing a dip **69**
Cornelia J Heyde, James M Scobbie, & Ian Finlayson

14. A thermoplastic head-probe stabilization device **72**
Anna Matosova, Lorenzo Spreafico, Alessandro Vietti, & Vincenzo Galatà

15. Ultrasound-integrated pronunciation teaching and learning **74**

Noriko Yamane, Jennifer Abel, Blake Allen, Strang Burton, Misuzu Kazama, Masaki Noguchi, Asami Tsuda, & Bryan Gick

16. Development of coarticulation in German children: Acoustic and articulatory locus equations **78**
Elina Rubertus, Dzhuma Abakarova, Mark Tiede, Jan Ries, & Aude Noiray

17. Development of coarticulation in German children: Mutual Information as a measure of coarticulation and invariance **79**
Dzhuma Abakarova, Khalil Iskarous, Elina Rubertus, Jan Ries, Mark Tiede, Aude Noiray

18. The articulation and acoustics of postvocalic liquids in the Volendam dialect **80**
Etske Ooijevaar

19. A method for automatically detecting problematic tongue traces **82**
Gus Hahn-Powell, Benjamin Martin, & Diana Archangeli

20. Word-final and word-initial glottalization in English-accented German: a work in progress **83**
Maria Paola Bissiri & Jim Scobbie

21. The production of English liquids by native Mandarin speakers **84**
Shuwen Chen, Xinran Ren, Richard Gananathan, Yanjiao Zhu, Sang-Im Kim, Peggy Mok

22. Examining tongue tip gestures with ultrasound: a literature review **86**
John M. Culnan

ULTRAFEST VII

8th–10th December 2015

FULL PROGRAMME

Tuesday, 8th December

9.00–9.45	Registration & Breakfast
9.45–10.00	Welcome Address from Derek Collins (*HKU Dean of Arts*)
10.00–11.00	**Oral Presentations (Session 1 Chair: Sang-Im Lee-Kim)**
10.00-10.30	*Alan Wrench* Applying a 3D biomechanical model to 2D ultrasound data
10:30-11.00	*Julián Villegas, Ian Wilson, Yuki Iguro, Donna Erickson* Effect of a fixed ultrasound probe on jaw movement during speech
11.00-11.30	Coffee & Tea
11.30-12.30	**Oral Presentations (Session 2 Chair: Celine Yueh-chin Chang)**
11.30-12.00	*Aude Noiray, Jan Ries, Mark Tiede* SOLLAR system: Sonographic & Optical Linguo-Labial Articulation Recording system
12.00-12.30	*Reza Falahati & Vahid Abolghasemi* Extraction of Persian coronal stops from ultrasound images using linear discriminant analysis
12.30-1.30	Lunch
1.30-2.30	**Keynote 1** (*Brad Story*)
2.00-3.00	**Discussion (Chair: Peggy Mok)**
3.00-3.30	Coffee & Tea
3.30-5.00	**Oral Presentations (Session 3 Chair: Rungpat Roengpitya)**
3.30-4.00	*Jonathan Yip, Diana Archangeli, Carol K.S. To* Development of lingual articulations among Cantonese-speaking children

4.00-4.30	*Stefan Frisch, Alissa Belmont, Karen Reddick, Nathan Maxfield* Speech stability, coarticulation, and speech errors in a large number of talkers
4.30-5.00	*Eleanor Lawson, James M. Scobbie, Jane Stuart-Smith* Using ultrasound tongue imaging to study the transfer of covert articulatory information in coda /r/
5.00-5.30	**General Discussion I (Chair: TBA)**

Wednesday 9th December

8.30-9.00	Breakfast
9.00-10.30	**Oral Presentations (Session 4 Chair: Cathryn Donohue)**
9.00-9.30	*Jonathan Yip* Coarticulatory effects on lingual articulations in the production of Cantonese syllable-final oral stops
9.30-10.00	*Suzy Ahn* The role of the tongue root in phonation of American English stops
10.00-10.30	*Robert Mailhammer, Mark Harvey, Tonya Agostini, Jason A. Shaw* Bolstering phonological fieldwork with ultrasound: Lenition and approximants in Iwaidja
10.30-11.00	Coffee & Tea
11.00-12.00	**Oral Presentations (Session 5 Chair: Alan Yu)**
11.00-11.30	*Amanda Miller* Timing of front and back releases in coronal click consonants
11.30-12.00	*Pertti Palo, Sonja Schaeffler, James M. Scobbie* Acoustic and articulatory speech reaction times with tongue ultrasound: What moves first?
12.00-12.30	**General Discussion II (Chair: Doug Whalen)**
12.30-1.30	Lunch
1.30-2.30	**Keynote 2** (*Patrick Wong*)
2.30-3.00	**Discussion (Chair: Carol K.S. To)**
3.00-3.15	Coffee & Tea

3.15-5.15	**Posters (and Coffee & Tea)**
5.15-6.00	Break
6.00-8.00	Dinner Reception (University Lodge)

Thursday 10th December

8.30-9.00	Breakfast
9.00-10.30	**Oral Presentations (Session 6 Chair: Feng-fan Hsieh)**
9.00-9.30	*Patrycja Strycharczuk, Koen Sebregts* /r/-allophony and gemination: An ultrasound study of gestural blending in Dutch
9.30-10.00	*Alessandro Vietti, Lorenzo Spreafico, Vincenzo Galatà* Allophonic variation: An articulatory perspective
10.00-10.30	*Noriko Yamane, Phil Howson* Ultrasound investigation of palatalized taps in Japanese
10.30-11.00	Coffee & Tea
11.00-12.30	**Oral Presentations (Session 7 Chair: Albert Lee)**
11.00-11.30	*Kevin Roon, Katherine Dawson, Mark Tiede, Douglas H. Whalen* Russian palatalization, tongue-shape complexity measures, and shape-based segment classification
11.30-12.00	*Douglas H. Whalen, Kevin Roon, Katherine Dawson, Mark Tiede* Exploring the relationship between tongue shape complexity and coarticulatory resistance
12.00-12.30	*Daniel Recasens, Clara Rodríguez* An investigation of lingual coarticulation resistance using ultrasound data
12.30-1.00	**General Discussion III (Chair: TBA)**
1.00-1.30	Break
1.30-2.30	Dim Sum Lunch (Victoria Harbour Restaurant 海港酒家–西寶城)

Optional Outing:

7.30-9.30	Dinner Buffet Cruise (Symphony of Lights, Harbour Cruise - Bauhinia 洋紫荊維港遊)

Poster Session (Wednesday 9th December, 3.15-5.15)

1 *Mai Ohkubo, James M. Scobbie*
Tongue shape dynamics in swallowing

2 *Marija Tabain, Richard Beare*
Recordings of Australian English and Central Arrernte using the EchoBlaster and AAA

3 *Paméla Trudeau-Fisette, Christine Turgeon, Marie Bellavance-Courtemanche, Lucie Ménard*
The effects of blindness on the development of lip and tongue movements in children

4 *Cheng Chen, Irene Ricci, Chiara Bertini, Reza Falahati, Chiara Celata*
An EPG and UTI investigation of syllable onsets and codas in Italian

5 *Sam Johnston*
A Kinect 2.0 system to track and correct head-to-probe misalignment

6 *Ian Wilson, Yuki Iguro, Julián Villegas*
Articulatory settings of Japanese-English bilinguals

7 *Joanne Cleland, James M. Scobbie, Zoe Roxburgh, Cornelia Heyde*
The UltraPhonix project: Ultrasound visual biofeedback for heterogeneous persistent speech sound disorders

8 *Joanne Cleland, James M. Scobbie, Jenny Isles, Kathleen Alexander*
Gradient acquisition of velars via ultrasound visual biofeedback therapy for persistent velar fronting

9 *Alessandro Vietti, Alessia Pini, Simone Vantini, Lorenzo Spreafico, Vincenzo Galatà*
A non-parametric approach to functional ultrasound data: A preliminary evaluation

10 *Lia Saki Bučar Shigemori, Marianne Pouplier, Štefan Beňuš*
Effects of phrasal accent on tongue movement in Slovak

11 *Mark Tiede, Douglas H. Whalen*
GetContours: An interactive tongue surface extraction tool

12 *Hannah King, Emmanuel Ferragne*
The feasibility of ultrasound imaging in the acquisition of English dark /l/ in French learners

13 *Cornelia Heyde, James M. Scobbie, Ian Finlayson*
Searching for closure: Seeing a dip

14 *Anna Matosova, Lorenzo Spreafico, Alessandro Vietti, Vincenzo Galatà*
A thermoplastic head-probe stabilization device

15 *Noriko Yamane, Jennifer Abel, Blake Allen, Strang Burton, Misuzu Kazama, Masaki Noguchi, Asami Tsuda, Bryan Gick*
Ultrasound-integrated pronunciation teaching and learning

16 *Elina Rubertus, Dzhuma Abakarova, Mark Tiede, Aude Noiray*
Development of coarticulation in German children: Acoustic and articulatory locus equations

17 *Dzhuma Abakarova, Khalil Iskarous, Elina Rubertus, Jan Ries, Mark Tiede, Aude Noiray*
Development of coarticulation in German children: Mutual Information as a measure of coarticulation

18 *Etske Ooijevaar*
The articulation and acoustics of postvocalic liquids in the Volendam dialect

19 *Gus Hahn-Powell, Benjamin Martin, Diana Archangeli*
A method for automatically detecting problematic tongue traces

20 *Maria Paola Bissiri, James M. Scobbie*
Word-final /r/ and word-initial glottalization in English-accented German: A work in progress

21 *Shuwen Chen, Xinran Ren, Richard Gananathan, Yanjiao Zhu, Sang-Im Kim, Peggy Mok*
The production of English liquids by native Mandarin speakers

22 *John Culnan*
Examining tongue tip gestures with ultrasound: A literature review

Oral presentations

PRESENTATION

Applying a 3D biomechanical model to 2D ultrasound data

Alan A Wrench[1,2] **Peter Balch[3]**

1. *Queen Margaret University 2. Articulate Instruments Ltd 3. Analogue Information Systems Ltd*

Abstract

A 3D biomechanical model of a tongue has been created using bespoke hexahedral manual mesh creating software. The software allows meshes to be created and vertices to be added, removed and moved in 3D, either individually or in selected groups. After a mesh has been digitally sculpted by hand, edges of the hexahedra can be assigned to "muscles". These "muscles" are controlled by manipulating the nominal length. A change in the nominal "muscle" length invokes Hooke's Law (modified so stiffness increases as the muscle contracts) to calculate the forces applied to every vertex in the mesh. Each vertex is moved iteratively until the forces on all vertices reach an equilibrium. The iterative calculation also includes a hydrostatic (volume preservation) component in the form of pressure force inside each hexahedron.

This equilibrium based approach has no temporal component so it cannot be used to predict movement. It does not explicitly model momentary imbalances in internal muscle forces which may occur during highly dynamic movement, although some implicit modelling may occur if it is not given time to iterate to equilibrium at a given time point. The big advantage of this technique over the more popular Finite Element Modelling approach is that it is flexible and stable. It does not lock up like Finite Element Models often do and is reasonably robust to arbitrary mesh design changes. Different shapes and muscle configurations can therefore be tested without worrying about the effect it may have on the stability of the modelling process.

A tongue mesh, once created, can be posed by contracting the assigned muscle groups. A midsagittal section of the 3D model can be superimposed on 2D midsagittal ultrasound data imported into the meshing software and the model then manually posed to fit each successive frame using landmarks on the ultrasound image as a guide. As the model is fitted to successive ultrasound frames (at 120fps), the patterns of "muscle" contraction over time are revealed. During the fitting process, choices in which muscles to contract can be influenced by attempting to avoid discontinuities in muscle contraction from frame to frame. This, in part, mitigates for any "many-to-one" muscle–to-shape mapping problem that may or may not exist. The result is a dynamic 3D model of tongue movement to match the 2D ultrasound data with associated muscle contraction time series generated as an important byproduct of the fitting process.

In this paper, validity of a given 3D tongue model is evaluated by comparing the predicted 3D tongue palate contact patterns with the actual patterns recorded by EPG. Results seem to indicate that, if the assumption of sagittal symmetry inherent in the present model is not too bold then the parasagittal shape of the tongue can be predicted from 2D midsagittal ultrasound data. Figure 2. Shows palate proximity patterns predicted by a model fitted to the midsagittal ultrasound of the sentence "The price range is smaller than any of us expected." The actual contact patterns measured by EPG at the same time points in the sentence, are similar if the asymmetries are ignored. This predictive ability is reasonable, within the terms dictated by symmetry, since the muscles which lie off the midline such as styloglossus, hyoglossus, transversus, inferior longitudinalis and verticalis all have an effect on the midsagittal tongue position and shape as well as forming an intrinsic part of the parasagittal lingual tissue.

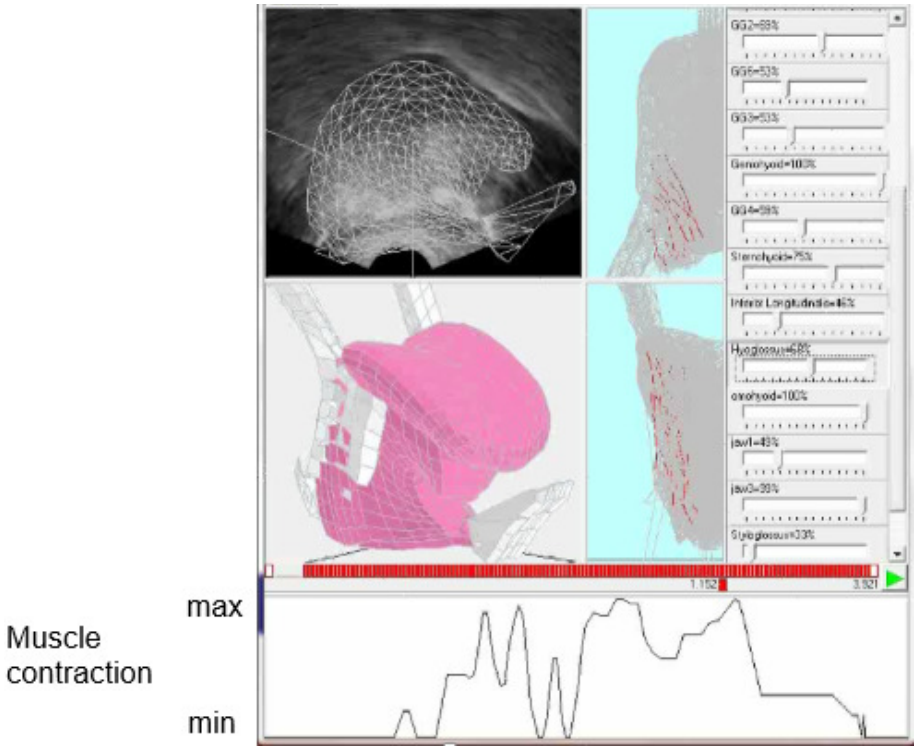


Figure 1. Top left: Single ultrasound frame with midsagittal section of 3D model superimposed. In this case the tip would be extended to fit the ultrasound image by relaxing the inferior longitudinalis and relaxing the anterior portion of the genioglossus. Middle left: The full 3D tongue shape. Right side: A set of sliders controlling each muscle. Bottom: The muscle contraction time series for the highlighted muscle (hyoglossus) Red bar: this is a series of approximately 400 ultrasound frames. Any or all frames can be selected and manually matched to the model. Unselected frames have nominal muscle lengths set to values linearly interpolated from neighbouring selected frame values.

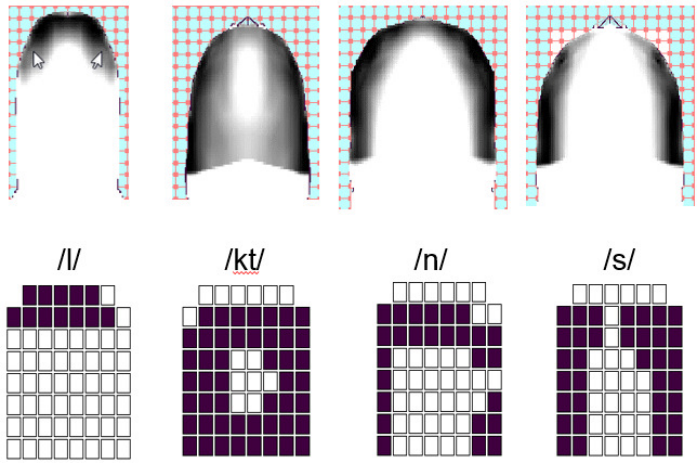


Figure 2. Top row: Shows distance from the model tongue to the model hard palate represented by greyscale where black is contact and white is ~1cm or greater. Bottom row: EPG patterns of the same segments from the same spoken sentence by the same speaker.

Effect of a fixed ultrasound probe on jaw movement during speech

Julián Villegas[1], Ian Wilson[1], Yuki Iguro[1], and Donna Erickson[2]
[1]University of Aizu, Japan, [2]Kanazawa Medical University, Japan

Abstract

The use of an ultrasound probe for observing tongue movements potentially modifies speech articulation in comparison with speech uttered without holding the probe under the jaw. To determine the extent of such modification, we analyzed jaw displacements of three Spanish speakers speaking with and without a mid-sagittal ultrasound probe. We found a small and not significant effect of the presence of the probe on jaw displacement. Counterintuitively, when speakers held the probe against their jaw larger displacements were found. This could be explained by a slight overcompensation on their speech production.

Method

We recorded three native speakers of Spanish uttering seven repetitions of 26 sentences (7 in English, 3 in Japanese, and 16 in Spanish) with and without the ultrasound probe fitted under their chin for a grand total of 1,092 sentences. For the statistical analysis, we used all sentences we recorded excepting those that had some capture (or trace extraction) problems. In total, 912 sentences were used in the analysis (252 in English, 107 in Japanese, 553 in Spanish).

Speakers

The three female speakers (s1, s2, and s3) were Salvadoran of 23, 28, and 34 years of age, with varying degree of second and third languages exposure: while the eldest reported ten years of English studies and three of Japanese (she had lived the last six years in Japan), s1 and s2 reported five and ten years of English training. None of these two had Japanese training. The youngest speaker had also lived in the USA for one year, immediately preceding the data collection whereas s2 had lived mainly in El Salvador. With the exception of s1, the speakers reported to still have a neutral Salvadoran Spanish accent, as acknowledged by their Salvadoran acquaintances and relatives.

Materials

The sentences were selected so the same vowel was prominently used in all the constituent words. These sentences are summarized in Appendix 1. A tripod-mounted Panasonic HDC-TM750 digital video camera was used to collect video of the front of the face. Light from two 300W halogen bulbs (LPL-L27432) was reflected onto the face to improve automatic marker tracking. Audio was recorded with a DPA 4080 miniature cardioid microphone connected to a Korg MR-1000 digital recorder, and tongue movements were recorded with an ultrasound Probe - Toshiba (PVQ-381A) connected to an ultrasound machine - Toshiba Famio 8 (SSA-530A).

Procedure

Speakers were recorded in two sessions: first without and second with an ultrasound probe under their chin. Each session was comprised of three blocks corresponding to the three languages recorded in this order: Spanish, English, and Japanese. Each block of utterances was randomly sorted and presented from a laptop computer located at about two meters in front of the speaker in Calibri black font (44 points) over white background. We prevented head tilting by changing the height of the display for each participant. Errors (mainly coughs, reading errors, and ultrasound probe misalignments) were marked visually and aurally in the video and audio recordings, prior to having the speaker repeat the dubious token.

Speakers were able to take short breaks between blocks and sessions. The two sessions were recorded in about one hour. Permission for performing these recordings was obtained following the University of Aizu ethics procedure.

After instructing the speakers about the experiment and querying them about their language background, they were asked to sit straight in a well-lit room, in front of a white background. The experimenters (two in each session) assisted them with putting on a lapel microphone. Subjects also don a glasses frame (without lenses) with a blue circle of about 8 mm in diameter, located at the center of the frame, above the participant's nose; a second marker was placed by the experimenters on the chin of the speaker and perpendicular to the frame line, as shown in Figure 1. Speakers were recorded in video at 29.97 frames per second (i.e., samples were taken every 33.367 ms) and at 44.1 kHz/16 bits in audio.

Figure 1. A speaker speaking without (left) and with the ultrasound probe (right). One marker (blue dot) was located on the lensless glass frame while the second marker was placed on the speaker's chin.

Post-processing

End points of each utterance were located from the audio of the video recordings in Praat [1] by visual inspection. These end-points were used to extract the videos using ffmpeg routines (https://ffmpeg.org). From the extracted videos, the blue dots were traced using the marker tracker program described in [2]. These trajectories were used to compute the Euclidian distance between the markers. Conversion from pixels to mm was approximated by measuring the physical frame (133 mm) and its corresponding videotaped counterpart (398 pixels).

Results

Each token was time normalized (*normT* – dividing each sample time by the length of the sentence) before fitting a smoothing cubic spline ANOVA (SSANOVA) model as implemented by Gu [3]. Note that this method has been successfully used in similar analyses such as F0 contours and larynx height for Mandarin tones [4] and the lingual and labial articulation of whistled fricatives [5].

In our model, jaw displacement (*distance*) is explained by the factors *Probe* (yes or no), *Sentence* (as in Appendix 1), *normT*, and the interaction between the two last factors. As a sole random factor we used *Speaker* (s01, s02, s03). We also used a Generalized Cross-Validation method for smoothing (as implemented in the SSANOVA library) with the default alpha value (i.e., $\alpha = 1.4$). The resulting model has an $R^2 = .484$, with no apparent redundancy on the fixed factors, and a relatively large variability explained by the random factor. This last

finding was expected since subjects had a large variability in jaw opening per sentence repetition (especially when speaking in unknown or poor proficiency languages).

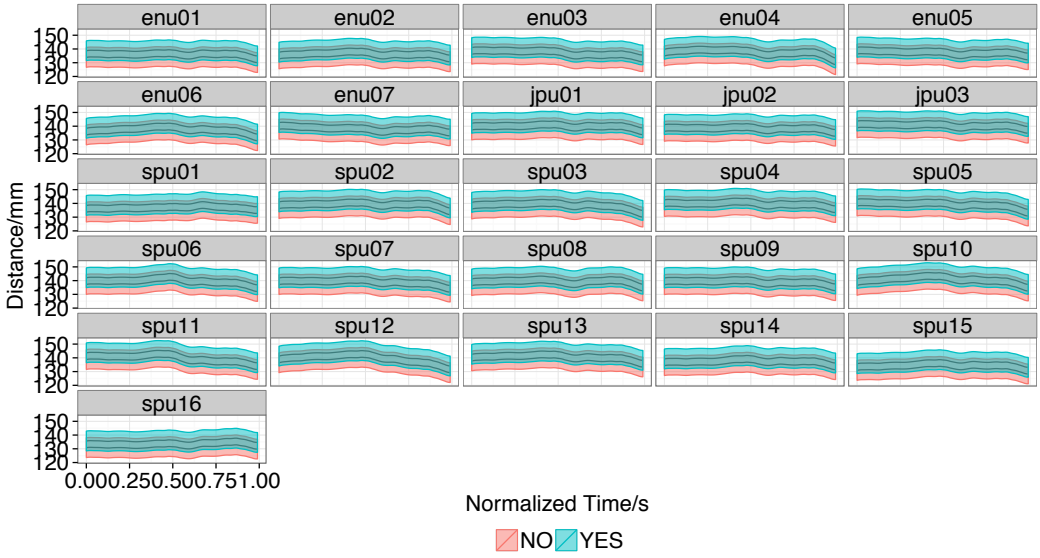


Figure 2. Time contour of the jaw opening for each of the studied sentences. Contours are plotted with their corresponding 95% Bayesian confidence intervals (CIs). Overlapping CIs suggest non-significant differences.

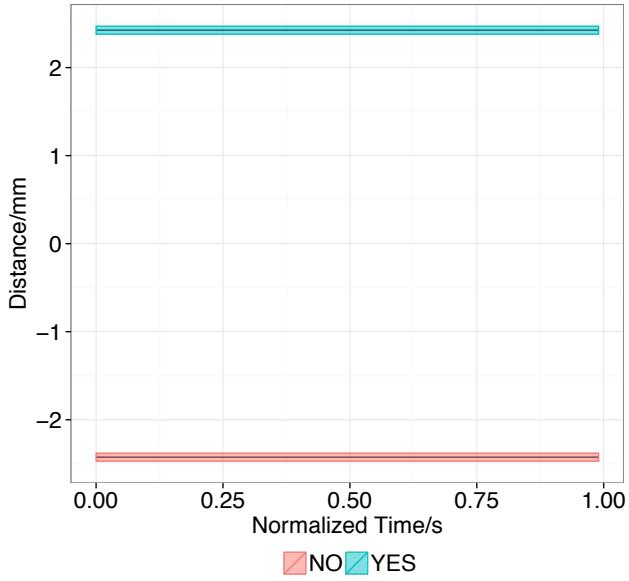


Figure 3. Distance difference predicted by the SSANOVA model for subjects holding the probe under their chin (YES) and when no probe was used (NO).

Findings

The resulting splines per sentence are presented in Figure 2. Interestingly, on average, subjects opened the jaw more when holding the probe under their chin than when they had no probe. When all sentences are considered, this difference was of about 5 mm as shown in Figure 3. The distance between markers varies with subject (i.e., larger subjects exhibit larger distances); in our case, speaker s2 had the smallest distances among the speakers; this is

reflected on the negative offset associated in the model (-5.976 compared to 0.0567 and 5.919 mm for speakers s1 and s3).

Conclusions

We did not find evidence supporting that the presence of an ultrasound probe located under the chin on the mid-sagittal plane hinders the jaw movement of the speakers. The small effect that we found was not significant and in opposition to the expected direction: i.e., it suggests that when the probe was present, subjects were opening the jaw more, probably as an overcompensation reaction.

Acknowledgements

This work was partially supported by the Japan Society for the Promotion of Science (JSPS), Grants-in-Aid for Scientific Research (C) #25370444.

References

[1] P. Boersma and D. Weenink. Praat. Available [Nov. 2015] from www.praat.org
[2] Barbosa, A. V., and Vatikiotis-Bateson, E. Video tracking of 2D face motion during speech. In *Signal Processing and Information Tech., IEEE International Symposium on* (pp. 791–796). (2006)
[3] Gu, C. (2014). Smoothing Spline ANOVA Models: R Package GSS. *J. of Statistical Software*, 58(5):1–25.
[4] Moisik, S., Lin, H., and Esling, J. (2013). *Larynx Height and Constriction in Mandarin Tones*, volume Eastward Flows the Great River: Festschrift in Honor of Professor William S-Y. Wang on his 80th Birthday, pages 187–205. City University of HK Press.
[5] Lee-Kim, S.-I., Kawahara, S., and Lee, S. J. (2014). The 'whistled' fricative in Xitsonga: its articulation and acoustics. *Phonetica*, 71(1):50–81.

Sonographic & Optical Linguo-Labial Articulation Recording system (SOLLAR)

Aude Noiray [a b], Jan Ries [a], Mark Tiede [b]
[a] *University of Potsdam,* [b] *Haskins Laboratories*

We present here a customized method developed jointly by scientists at LOLA (Potsdam University) and Haskins Laboratories (New Haven) for the recording of both tongue and lip motion during speech tasks in young children. The method is currently being used to investigate the development of 1) coarticulation (resistance and anticipatory coarticulation, cf. two other abstracts submitted); and 2) articulatory coordination in preschoolers compared with adults who have mature control of their speech production system.

Children are recorded with a portable ultrasound system (Sonosite Edge, 48Hz) with a small probe fixed on a custom-made probe holder and ultrasound stand. The probe holder was specifically designed to allow for natural vertical motion of the jaw but prevent motion in the lateral and horizontal translations. The set up is integrated into a child-friendly booth that facilitates integrating the production tasks into games.
Ultrasound video data are collected concurrently with synchronized audio recorded via a microphone (Shure, 48kHz,), pre-amplified before being recorded onto a desktop computer. In addition to tongue motion, a frontal video recording of the face is obtained with a camcorder (Sony HDR-CX740VE, fps: 50Hz). This video is used to track lip motion for subsequent labial measurements, and to track head and probe motion for transforming contours extracted from the ultrasound images to a head-based coordinate system. The speech signal is also recorded via the built-in camcorder microphone, and synchronization of both video signals (from the ultrasound and the camcorder) is performed through audio cross-correlation in post-processing.

Lip motion is characterized with a video shape tracking system (Lallouache 1991) previously used for examining anticipatory coarticulation in adults (Noiray et al., 2011) and children (Noiray et al., 2004; 2008). During production tasks, the lips of our young participants are painted in blue as this color maximized contrast with the skin. In post-processing these blue shapes are then tracked for calculation of lip aperture, interolabial area and upper lip protrusion.
Tongue contours derived from ultrasound are relative to the orientation of the probe with respect to the tongue surface. To correct for jaw displacement and (pitch) rotation of the head we compute two correction signals similar to the HOCUS method described in Whalen et al. (2005), but in this case derived from tracking the positions of blue reference dots in the video signal using custom Matlab procedures. The displacement of the probe relative to the centroid of dots placed on each speaker's forehead provides a vertical correction signal. The orientation of dots placed on the cheek observed within the video image through a mirror oriented at 45° giving a profile view provides a pitch rotation correction signal around the lateral axis. Application of these two signals to the extracted contours allows for their consistent comparison in a head-centric coordinate system.

Acknowledgments

This work is supported by the German DFG GZ: NO 1098/2-1.

References

Lallouache, M. T. (1991). Un poste «Visage-parole» couleur. Acquisition et traitement automatique des contours des lèvres (A «face-speech» interface. Automatic acquisition and processing of labial contours), Ph.D. ENSERG, Grenoble, France.

Noiray A Ménard L Cathiard MA Abry C and Savariaux C (2004). The development of anticipatory labial coarticulation in French: A pioneering study. In *Proceedings of Interspeech, 8th ICSLP*, 53-56.

Noiray, A., Cathiard, M. A., Me´nard, L., and Abry. C. (2008). Emergence of a vowel

gesture control. Attunement of the anticipatory rounding temporal pattern in French children. in Emergence of Language Abilities, edited by S. Kern, F. Gayraud, and E. Marsico (Cambridge Scholars Publishing, Newcastle, UK), pp. 100–116.

Noiray A., Cathiard M-A., Ménard L., & Abry C. (2011). Test of the Movement Expansion Model: Anticipatory vowel lip protrusion and constriction in French and English speakers. Journal of Acoustical Society of America, 129 (1), 340-349.

Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E., (2005). HOCUS, the Haskins Optically-Corrected Ultrasound System. Journal of Speech, Language, and Hearing Research , 48, 543-553.

Extraction of Persian coronal stops from ultrasound images using linear discriminant analysis

Reza Falahati
Scuola Normale Superiore di Pisa

Vahid Abolghasemi
University of Shahrood

Introduction

Ultrasound is an appealing technology which could be used for imaging the vocal tract. Similar to other techniques, it also has some limitations. Tracing the tongue contours in ultrasound images is a very time consuming task. Thirty minutes of tongue imaging at 60 fps will result in 108,000 images. Several different approaches to this problem have been proposed (Angul & Kambhamettu 2003; Baker 2005; Li et al. 2006; Fasel and Berry 2010; Tang et al. 2012; Hueber 2013; Pouplier & Hoole, 2013; Sung et al. 2013), with promising results. This study uses a new application called TRACTUS (Temporally Resolved Articulatory Configuration Tracking of UltraSound) developed by Carignan (2014) for extracting time-varying articulatory signals in large-scale image sets and compares the results with Falahati (2013) who has manually traced the tongue contours. The research question followed here is whether the automatic tracing can capture the articulatory differences between the simplified and unsimplified consonant clusters in Persian. The consonant clusters under study are composed of coronal stops [t d] followed and preceded by non-coronal consonants (i.e., $V_1C_1C_2\#C_3V_2$) where target coronal stops (i.e., C_2) could be optionally simplified.

Methodology

In order to choose the ultrasound images for processing, TextGrid was used to mark the target coronal consonants as well as the preceding and following consonants (i.e., C_1 & C_3) and also the two vowels adjacent to the three consonants in the middle. After choosing the images of interest, a feature reduction/extraction technique was applied. The open source software suite TRACTUS implemented in MATLAB was used for such end. The first step was to specify the border of ultrasound fan within the images followed by filtering. The goal at this stage was to strike a balance between the tongue contours and image noise (see Figure 1 top left). Choosing the region of interest (i.e., ROI) was the next step in the process. At this time the area of image showing the range of tongue contour movement was created (see Figure 1 top right). The final stage in using TRACTUS was to generate PC scores. This was the result of applying principle component analysis (PCA, Hueber et al. 2007) to the processed data. PC scores represent "the degree to which the imaged vocal tract matches a limited set of articulatory configurations which are identified by the PCA model" (Carignan & Mielke, p. 4). The combinations of PCs result in heatmaps illustrating the means (see Figure 1 bottom).

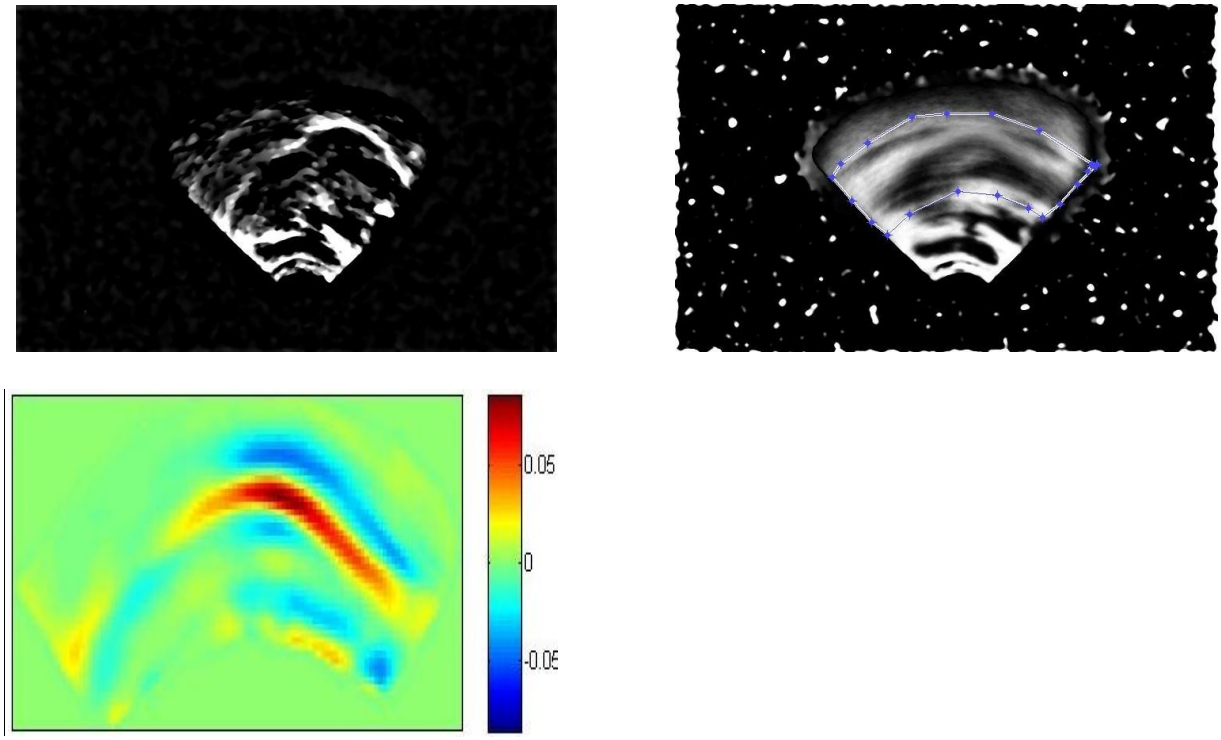


Figure 1: Top left: filtered image; Top right: polygonal ROI; Bottom: Heatmap for PC1.

TRACTUS tool is helpful up to the point of creating the PC scores from the ultrasound data. Once the PC scores were created, they were transformed via liner discriminant analysis (i.e., LDA) to create signals as inputs to an LDA model with classes for the simplified and unsimplified coronal stops [t d] as well as the remaining sounds. The articulatory signals generated for individual tokens in our study are analogous to tracing one specific point of the tongue over temporal dimension to generate gestural scores (Falahati 2013; Pouplier & Hoole, 2013; Carignan & Mielke 2014).

Results

The research question followed in this study was whether the articulatory signals generated for coronal stops [t d] could distinguish between the tokens with simplified and unsimplified consonant clusters and whether the result was comparable to Falahati (2013). The preliminary results of this study for one subject shows that this method is quite successful for teasing apart the tokens with full alveolar gestures versus the ones which lack it. The results of these token frames traced manually in Falahati (2013) supports the results for LDA class scores here. Figure 2 below illustrates a representative number of tokens with and without coronal gestures.

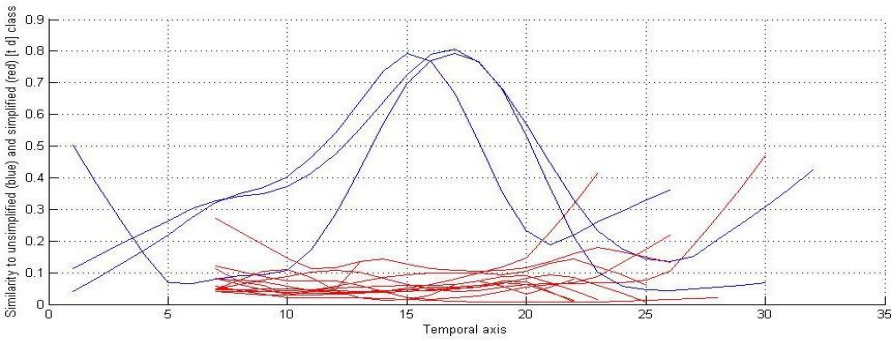


Figure 2: The LDA class scores over time. Tokens with unsimplified coronal stops (blue); Tokens with simplified coronal stops (red).

References

Baker, A. 2005. *Palatoglossatron 1.0*. University of Arizona, Tucson, Arizona. http://dingo.sbs.arizona.edu/~apilab/pdfs/pgman.pdf.

Carignan, C. 2014. TRACTUS (Temporally Resolved Articulatory Configuration Tracking of UltraSound) software suite. *URL: http://phon.chass.ncsu.edu/tractus*.

Carignan, C., & Mielke, J. 2014. Extracting articulatory signals from lingual ultrasound video using principal component analysis. MS.

Falahati, R. 2013. *Gradient and Categorical Consonant Cluster Simplification in Persian: An Ultrasound and Acoustic Study*. Ph.D dissertation, University of Ottawa.

Fasel, I. and Berry, J. 2010. Deep belief networks for real-time extraction of tongue contours from ultrasound during speech. *In Proceedings of the 20th International Conference on Pattern Recognition*, pp. 1493–1496.

Hueber, T., Aversano, G., Chollet, G., Denby, B., Dreyfus, G., Oussar, Y., Roussel, P., and Stone, M. 2007. Eigen tongue feature extraction for an ultrasound-based silent speech interface. In *Proceedings of 2007 International Conference on Acoustics, Speech, and Signal Processing*, pp. 1245–1248.

Hueber, T. 2013. Ultraspeech tools: Acquisition, processing and visualization of ultrasound speech data for phonetics and speech therapy. *In Proceedings of Ultrafest VI Conference, pp.* 10-11.

Li, M., Kambhamettu, C., and Stone, M. 2005. Automatic contours tracking in ultrasound images. *Clinical Linguistics and Phonetics*, 19:545–554.

Pouplier, M. and Hoole, P. 2013. Comparing principal component analysis of ultrasound images with contour analyses in a study of tongue body control during German coronals. *In Proceedings of Ultrafest VI Conference,* pp. 25-26.

Sung, J. H., Berry, J., Cooper, M., Hahn-powell, G., and Archangeli, D. 2013. Testing AutoTrace: A machine learning approach to automated tongue contour data extraction. In Proceedings of Ultrafest VI Conference, pages 9-10.

Tang, L., Bressmann, T., and Hamarneh, G. 2012. Tongue contour tracking in dynamic ultrasound via higher order MRFs and efficient fusion moves. *Medical Image Analysis*, 16:1503–1520.

Keynote 1: Tuesday, December 8, 1:30-2:30pm

Brad Story

The University of Arizona

Acoustic sensitivity of the vocal tract as a guide to understanding articulation

Understanding the relation of speech articulation and the acoustic characteristics of speech has been goal of research in phonetics and speech science for many years. One method of studying this relation is with acoustic sensitivity functions that, when calculated for a specific vocal tract configuration, can be used to predict the direction in which the resonance frequencies (formants) will shift in response to a perturbation of the vocal tract shape. Projected onto the anatomical configuration of the articulators, the sensitivity functions provide a means of generating hypotheses concerning why articulatory movements are executed in both canonical and idiosyncratic patterns. This talk will summarize some recent efforts to investigate the relation of articulation and acoustics by means of sensitivity functions, vocal tract modeling, simulation of speech, and kinematic analysis based on articulography.

[Supported by NIH R01-DC011275 and NSF BCS-1145011].

Development of lingual articulations among Cantonese-speaking children

Jonathan Yip[1], Diana Archangeli[1,2], and Carol K.S. To[1]
University of Hong Kong[1], University of Arizona[2]

Introduction

The vocal tract undergoes substantial physical change from early childhood into late childhood, and it is a commonly held belief that many of the speech production issues that appear during the beginning of elementary school are simply a continuation of earlier speech behaviors rather than novel, atypical behaviors. In this paper, we examine the development of lingual articulation as Cantonese-speaking children mature from a young age toward adulthood, with the question as to whether speech production issues during later childhood are indeed a continuation of speech production patterns in early childhood. Developing children may struggle to produce adult-like speech sounds when the proportional sizes of their speech organs differ from those of adults (McGowan, personal communication). To do so, we use ultrasonic tongue imaging to examine the shape of the tongue during the articulation of lingual consonant sounds known to be acoustically interchangeable among younger Cantonese-acquiring children but typically acoustically distinct by elementary-school age (To *et al.*, 2013). The consonantal contrasts of interest are:

- Alveolar stops [t, tʰ] vs. velar stops [k, kʰ] (typically adult-like by age 3;6)
- Alveolar lateral [l] vs. central palatal [j] (typically adult-like by age 4;0)
- Apical affricates [ts, tsʰ] vs. laminal fricative [s] (typically adult-like by age 4;6)

Methodology

In our study, we collected ultrasonic images during these lingual consonants spoken by participants belonging to 3 age categories: 7 younger children (2;6 to 4;6), 8 older children (4;7 to 9;0), and 8 adults (18 or older). The general articulatory ability of each child was assessed using the Hong Kong Cantonese Articulation Test (HKCAT) (Cheung *et al.*, 2006). The HKCAT contains 91 test sounds (48 onsets, 29 vowels, 16 codas) elicited through pictured words and transcribed by researchers with phonetic training. Target words were monosyllables beginning with each sound of interest, followed by the rime [aː] or [ɐm]. There were 9 Cantonese target words in total. Children were prompted to say each item in a picture-naming task, and adults received items prompts in Chinese orthography. Children produced up to 5 repetitions of each item and adults produced 6 iterations of each item. Head-to-probe stabilization was achieved with 3 fully articulating camera/lighting arms: 2 arms provided resting points for talkers' foreheads and 1 arm held the transducer in a fixed position. During scanning, best efforts were made to ensure that each talker's head did not move relative to the probe. Image frames of interest were determined through the acoustic recordings and lingual contours within frames were extracted with EdgeTrak (Li, *et al.*, 2005).

In order to assess the degree of articulatory place contrast within each talker's productions, the angle of maximal constriction along the lingual contour during the production of each sound was measured, where maximal constriction was defined as the point along the contour at which minimum aperture distance to the hard palate contour (as ascertained from video images of water boluses) occurred during the interval of articulatory achievement. Angles were taken from a reference angle of 0°. Examples of this measurement are shown in Figure 1. As a result of this measurement procedure, place contrasts should involve larger angles of constriction for dorsal sounds [k, kʰ] than for coronal sounds [t, tʰ]. Angles were then converted into *z*-scores in order to allow for comparisons between talkers, who possess varying vocal tract shapes and sizes.

Results & Discussion

Data indicate that all adult talkers and most older and younger children produced the target consonant sounds with the expected relative constriction angle. However, 6 out of the 15 children (3 younger, 3 older) frequently articulated alveolar sounds in the dorsal region

(examples in Figure 2). These results correspond relatively well with each child's HKCAT scores. For 5 of these children, both alveolar and velar sounds were produced with a wide degree of variation in terms of where constrictions were formed, suggesting that these children have not yet identified specific locations along the palate where these sounds should be articulated. The remaining talker (CC05: age 5;6) consistently articulated the alveolar stops [t, t^h] nearly identically to the velar stops [k, k^h], despite the fact that the productions of coronal sounds [ts, ts^h, s, l] were articulated with tongue-tip raising toward the dento-alveolar region. This finding indicates that, while younger talkers (below 4;6) may not yet have mastered the contrast between coronal and dorsal articulations, sometimes even executing dorsal and apical raising gestures simultaneously, older children who have persistent articulatory issues, such as CC05, may have settled on consistent, although mismatched, articulations for their consonant productions during early childhood.

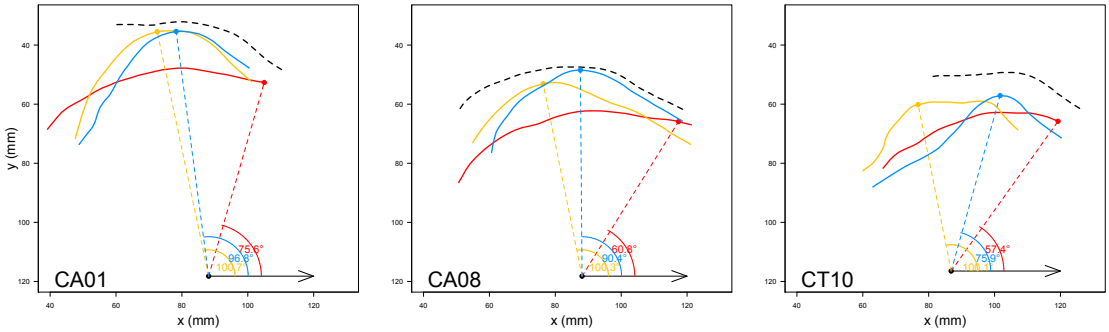


Figure 1. Examples measures of *angle of maximal constriction* for lingual contours during velar [k] (yellow), palatal [j] (blue), and alveolar [t] (red) gestures, as produced by two adult talkers (CA01, CA08) and one younger child (CT10: age 4;4).

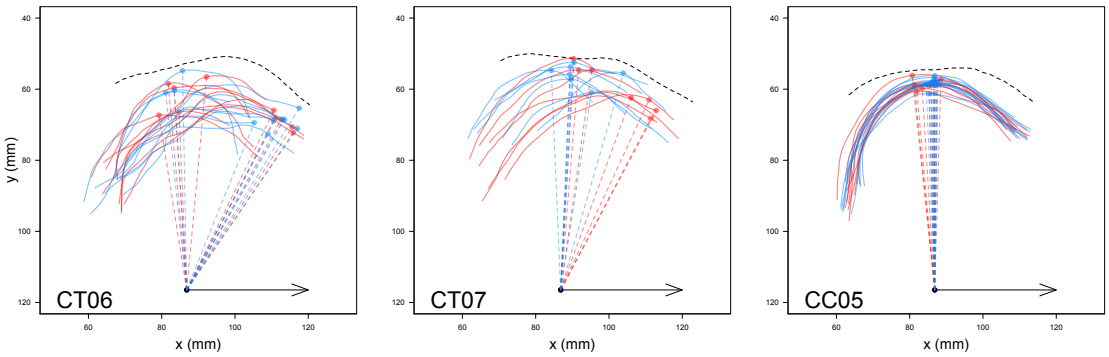


Figure 2. Lingual contours (and constriction angles) from alveolar stops [t, t^h] (red) and velar stops [k, k^h] (blue), produced by child talkers with the lowest three HKCAT scores: CT06 (age 3;10, HKCAT score 56.0%), CT07 (age 3;5, HKCAT score 68.1%), and CC05 (age 5;6, HKCAT score: 74.7%). For the sake of clarity, lingual contours from other target sounds are not pictured.

References

Cheung, P., Ng, A., and To, C. K. S. 2006. *Hong Kong Cantonese Articulation Test*. City University of Hong Kong, Hong Kong.

Li, M., Kambhamettu, C., and Stone, M. 2005. Automatic contour tracking in ultrasound images. *Clinical Linguistics and Phonetics*, **19**(6-7), 545–554.

To, C. K. S., Cheung, P., and McLeod, S. 2013. A population study of children's acquisition of Hong Kong Cantonese consonants, vowels, and tones. *Journal of Speech, Language, & Hearing Research*, **56**(1), 103–122.

Speech stability, coarticulation, and speech errors in a large number of talkers

Stefan A. Frisch, Alissa J. Belmont, Karen Reddick, Nathan D. Maxfield
Department of Communication Sciences and Disorders, University of South Florida

Introduction

This study uses ultrasound to image onset lingual stop consonant articulation in words. In one set of stimuli, velar stop consonants are produced in variety of vowel contexts. Anticipatory coarticulation can be interpreted as a quantitative measure indicating the maturity of the speech motor system and its planning abilities (Zharkova, Hewlett, & Hardcastle, 2011, *Motor Control*, 15, 118-140). Part of the method for measuring anticipatory coarticulation in Zharkova et al (2011) involves measuring multiple repetitions of the same item. Variation in these repetitions is taken to be an index of motor speech stability. Speech motor stability can also be examined through challenging speech production tasks such as tongue twisters. The present study examines coarticulation and speech stability in typical speakers and people who stutter across three lifespan age groups.

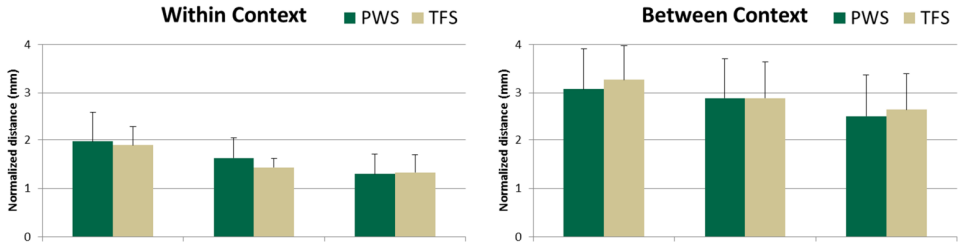
Methods

One hundred twenty two (n = 122) participants were recruited in three age groups over the lifespan (8-12yo; 18-30yo; 55-65yo) who were either typically developing speakers (n = 73) or people who stutter (n = 49). Individual age and talker group combinations varied in size from 11 to 29 talkers. Articulate Assistant Advanced 2.0 software was used to semi-automatically generate midsagittal tongue contours at the point of maximum stop closure and was used to fit each contour to a curved spline. Three measures of articulatory ability are being examined based on curve-to-curve distance (Zharkova et al 2011). Token-to-token variability is examined from multiple velar vowel productions within the same vowel context, describing the accuracy of control, or stability, of velar closure gestures. Variability in production between vowel contexts is an index of coarticulation as in Zharkova et al (2011). Participants produced 18 target words in a frame sentence for the coarticulation part of the study (e.g. *Say a key again*). Participants also produced 16 four-word tongue twisters varying in alveolar and velar stop onset with low vowels (e.g. *top cap cop tab*). The use of curve-to-curve distance has been extended in this study to cases of tongue twisters as a measure of similarity of the production to typical targets for both the intended and error category following Reddick & Frisch (ICPhS poster, August 2015).

Results

Completed results indicate an overall age effect, interpreted as refinement of speech motor production, with increased speech stability and progressively more segmental (less coarticulated) productions across the lifespan (Figure 1). Anticipatory coarticulation can be interpreted as a quantitative measure indicating the maturity of the speech motor system and its planning abilities (Zharkova et al 2011). A tendency toward decreased stability was found for younger people who stutter, but this difference was small and absent among older adults (Belmont, unpublished MS thesis, June 2015). Classification of speech errors is still ongoing, but partial data analysis finds a correlation between speech motor stability and the rate of production of both gradient and perceived speech errors in tongue twisters replicating Reddick & Frisch (2015).

Figure 1: Speech stability (left, within context distance) and coarticulation (right, between context distance) for Children, Young Adults, and Older Adults with (PWS) and without (TFS) stuttering.



Using ultrasound tongue imaging to study the transfer of covert articulatory information in coda /r/

Eleanor Lawson[1], James M. Scobbie[1], Jane Stuart-Smith[2].

[1] Queen Margaret University, Edinburgh *[2] University of Glasgow*

Several decades of investigation have established that there is an auditory dichotomy for postvocalic /r/ in the Scottish Central Belt, (Romaine 1978; Speitel and Johnston 1983; Stuart-Smith 2003; Stuart-Smith 2007) and beyond, e.g. in Ayrshire (Jauriberry, Sock et al. 2012). Weak rhoticity is a feature of working-class (WC) speech, strong rhoticity is associated with middle-class (MC) Central Belt speech.

Ultrasound tongue imaging (UTI) has identified articulatory variation that contributes to this auditory dichotomy; underlyingly coda /r/ in MC and WC speech involves radically different tongue shapes (Lawson, Scobbie et al. 2011b) and tongue gesture timings (Lawson, Scobbie and Stuart-Smith 2015). This articulatory variation has gone unidentified, despite decades of auditory and acoustic analysis (Romaine 1978; Speitel, Johnston 1983; Stuart-Smith 2003; Stuart-Smith, Timmins et al. 2007). UTI revealed that bunched /r/ variants (see Delattre & Freeman 2009) are prevalent in WC speech (Lawson, Scobbie and Stuart-Smith 2014) while WC speech shows a prevalence of tongue tip/front-raised /r/ with delayed anterior gestural maxima that can occur after the offset of voicing or during the articulation of a following labial consonant, e.g. in *perm, firm, verb* etc. The fact that apparently covert articulatory variants pattern with speaker social class, suggests that this covert articulatory variation in /r/ production is perceptible or recoverable.

We present results of a UTI-based speech-mimicry study that investigates whether these types of subtle articulatory variation can be copied if the speaker is presented with audio only and asked to mimic what they hear. We investigate whether they use different articulatory strategies to achieve the strong rhotic quality found in MC /r/ by e.g. by either bunching or retroflexing their tongue, and whether they misinterpret delayed, weakly audible, /r/ gestures as deletion of /r/.

We recruited thirteen female Central-Belt Scottish speakers to take part in the mimicry study (8 MC aged 13-23 and 5 WC aged 13-22), as females were found to produce the most extreme articulatory variants in their social-class groups (see Lawson et al 2014). Baseline articulatory information on their /r/ production was gathered from audio-ultrasound word-list recordings containing 23 (C)Vr and (C)VrC words such as *pore, farm, ear, herb* etc., plus 55 distractors. All MC participants used bunched /r/ variants in baseline condition. All WC participants used variants that involved raising the tongue front or tip in baseline condition.

Audio stimuli were 82 nonsense-words extracted from the female-speech section of an audio-ultrasound corpus of adolescent speech, collected in Glasgow in 2012. Nonsense words were used to avoid speakers normalizing towards their habitual production of a word. There were 24 /r/-ful nonsense words, randomized in the audio stimuli: (Mimic A) 12 with front/tip-up /r/s with a delayed /r/ gesture and (Mimic B) 12 with bunched /r/ with an early /r/ gesture. The rest of the stimuli (58 tokens) were distractors. Intensity of the audio stimuli was scaled to a mean 70dB using Praat (Boersma & Weenink 2013). Participants were asked to mimic the audio stimuli as closely as possible, "as if they were an echo".

Analysis showed a range of /r/-mimicking behaviours, the most common of which were (a) no modification of tongue shape from the baseline to the mimicry conditions and (b) modification from the speaker's baseline /r/ (i.e. tip up to bunched, or bunched to tip up), but no differentiation between the tongue shape used in the Mimic A and Mimic B conditions. (c) Two of the participants successfully copied the underlying tongue shapes of the audio stimuli on a token by token basis with high levels of accuracy, resulting in distinct tongue shapes for the Mimic A and Mimic B conditions. Participants who used tip up /r/ in baseline did not attempt to mimic bunched /r/ stimuli by retroflexing their tongues, suggesting that the underlying bunched /r/ is perceptible and distinguishable from a retroflex. A small number of weakly /r/-ful stimuli were mimicked with no /r/ gesture by WC speakers in the study, but in most cases, speakers

produced an /r/ gesture when they mimicked weakly /r/-ful audio stimuli, which suggests that cues indicating rhoticity persist in the audio signal (see also Lennon 2013).

References

BOERSMA, PAUL & WEENINK, DAVID, 2013. *Praat: doing phonetics by computer.* 5.3.47 edn. http://www.praat.org/:

DELATTRE, PIERRE & FREEMAN, DONALD C., 2009. A dialect study of American r's by x-ray motion picture. *Linguistics,* **6**(44), pp. 29-68.

JAURIBERRY, T., SOCK, R., HAMM, A. and PUKLI, M., 2012. Rhoticite et derhoticisation en anglais ecossais d'Ayrshire, *Proceedings of the Joint Conference JEP-TALN-RECITAL*, June 2012 2012, ATALA/AFCP, pp. 89-96.

LAWSON, E., SCOBBIE, J. M. and STUART-SMITH, J. (2015). The role of anterior lingual gesture delay in coda /r/ lenition: an ultrasound tongue imaging study, *Proceedings of the 18th international congress of phonetic sciences,* 10th - 14th August 2015. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0332.pdf

LAWSON, E., SCOBBIE, J.M. and STUART-SMITH, J., 2011a. A single-case study of articulatory adaptation during acoustic mimicry, *Proceedings of the 17th international congress of phonetic sciences*, 17th - 21st August 2011a, pp. 1170-1173.

LAWSON, E., SCOBBIE, J.M. and STUART-SMITH, J., 2011b. The social stratification of tongue shape for postvocalic /r/ in Scottish English1. *Journal of Sociolinguistics*, **15** (2), pp. 256-268.

LENNON, R. 2013. The effect of experience in cross-dialect perception: Parsing /r/ in Glaswegian. *Unpublished dissertation submitted for the degree of Master of Science in English Language and Linguistics in the School of Critical Studies. University of Glasgow.*

MACAFEE, C., 1983. *Glasgow. Varieties of English Around the World. Text series T3.* Amsterdam: Benjamins.

ROMAINE, S., 1978. Postvocalic /r/ in Scottish English: Sound change in progress. In: P. TRUDGILL, ed, *Sociolinguistic Patterns in British English.* pp. 144-158.

SPEITEL, H.H. and JOHNSTON, P.A., 1983. *A Sociolinguistic Investigation of Edinburgh Speech.* End of Grant Report. Economic and Social Research Council.

STUART-SMITH, J., 2007. A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents, TROUVAIN, J., BARRY, W.J., ed. In: *Proceedings of the 16th International Congress of Phonetic Sciences*, 6 - 10 August 2007 2007, Universitat des Saarlandes, pp. 1307.

STUART-SMITH, J., 2003. The phonology of modern urban Scots. In: J. CORBETT, J.D. MCCLURE and J. STUART-SMITH, eds, *The Edinburgh Companion to Scots.* 1 edn. Edinburgh, U.K.: Edinburgh University Press, pp. 110-137.

STUART-SMITH, J., TIMMINS, C. and TWEEDIE, F., 2007. Talkin' Jockney: Variation and change in Glaswegian accent1. *Journal of Sociolinguistics,* **11**(2), pp. 221-260.

Coarticulatory effects on lingual articulations in the production of Cantonese syllable-final oral stops

Jonathan Yip
University of Hong Kong

Introduction

Previous studies have determined that the inaudibly-released syllable-final oral stops [pˀ tˀ kˀ] of Cantonese are primarily cued by spectral formant transitions into stop closure during the preceding vowel (Ciocca et al., 1994; Khouw & Ciocca, 2006). However, younger speakers are reported to have a tendency to either merge the alveolar and velar codas [tˀ] and [kˀ] or produce a full glottal closure in lieu of or immediately preceding the coda gesture (Zee, 1999; Law et al., 2001), potentially leading to perceptual confusions between alveolar and velar stop place. While prior perceptual work has attributed this phenomenon to the phonological loss of a coronal-dorsal coda place contrast in younger speakers, articulatory investigations of the loss of [tˀ]–[kˀ] contrasts for this segment of the population are lacking. The goal of this study is to understand whether young-adult speakers consistently produce lingual gestures that correspond to the coda stops [tˀ] and [kˀ] and whether there are strong anticipatory coarticulatory influences that could mask the acoustic cues to coda place according to the place of the following consonantal gesture.

Methodology

In this study, ultrasonic tongue imaging was used to examine the lingual dynamics during the production of coda stops [tˀ, kˀ] in 24 Cantonese disyllabic words. These target words were selected such that the initial syllables containing the coda stops were one of 4 morphemes (發 [faːt^{3}], 法 [faːt^{3}], 白 [paːk^{2}], and 拍 [pʰaːk^{3}]) and the second syllables contained onset consonants with labial, coronal, and dorsal place of articulation, e.g. [faːt^{3}**m̲**ɐn^{21}] *vs.* [faːt^{3}**t̲**aːt^{2}] *vs.* [faːt^{3}**k̲**ɔk^{3}] and [paːk^{2}**p̲**aːn^{25}] *vs.* [paːk^{2}**t̲**ɐu^{25}] *vs.* [paːk^{2}**k̲**aːp^{25}]. Ultrasonic images of the productions of 5 native speakers of the Hong Kong variety of Cantonese were collected using a Telemed ClarUs machine at a frame rate of 60 fps and sequences of ultrasonic frames were extracted during the interval […$V_1C_1.C_2V_2$…] within each target item, as determined from the synchronized acoustic signal. Splines corresponding to lingual contours in each frame within the intervals of interest were traced and extracted in EdgeTrak (Li et al., 2005), as well as contours of the palate. To assess the achievement of syllable-final stop gestures, minimum values of distance between the tongue contour and coronal and dorsal regions of the palate ("aperture") during C_1-C_2 closure were calculated at each frame time. For each talker, minimum aperture distances corresponding to the coda gesture were compared in a linear mixed-effects model with fixed effects of *articulator* (tongue tip, tongue dorsum) and *place context* (labial, coronal, dorsal) and the random effect of *item*.

Results & Discussion

The data reveal that the 5 speakers' productions fell into three general categories of articulatory patterns: gestural preservation (S5), gestural reduction or partial loss (S1 and S2), and near-complete loss (S3 and S4). In the preservation pattern, lingual articulations consistently achieved full stop closures near the end of the first vowel, regardless of articulator and place context. In the reduction/partial loss pattern, lingual articulations were greatly reduced in labial contexts but involved strong effects of tongue-tip to tongue–dorsum coproduction in lingual-lingual sequences (t+DORSAL and k+CORONAL). For talkers exhibiting nearly complete loss of the syllable-final stop articulations, movements of the tongue during the C_1-C_2 closure interval corresponded strongly to the place of the following onset consonant only, with little evidence of lingual coproduction behaviors. Differences in speech rate are also observed and could be the source of gestural timing variation between speakers. The articulation of Cantonese syllable-final stops were varied—not only between

talkers but also before syllables differing in onset place—and this variability even occurred within the same morpheme (Chinese character) in different contexts. The results of this study provide a richer picture as to whether and how the inaudibly-released, syllable-final, lingual oral stops in Cantonese are produced by young-adult talkers.

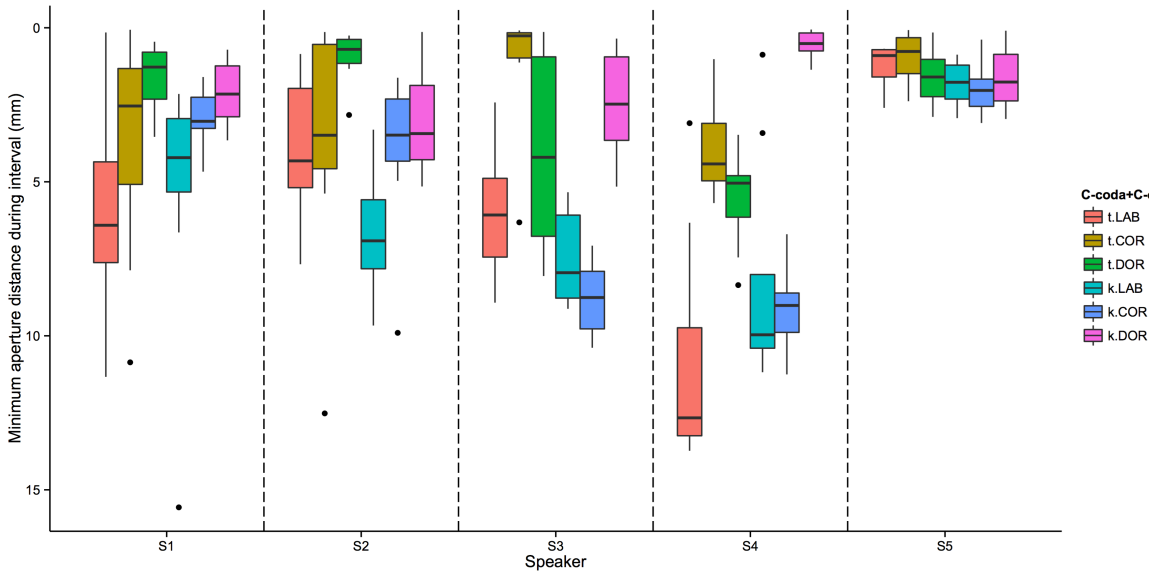


Figure 1. Boxplot of minimum aperture distances during the C_1-C_2 closure interval for codas [t] and [k] (C_1) in labial, coronal, and dorsal onset (C_2) contexts, grouped by speaker.

References

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

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**, 172–185.

Law, S.-P., Fung, R. S.-Y., & Bauer, R. 2001. Perception and production of Cantonese consonant endings. *Asia Pacific Journal of Speech, Language and Hearing*, **6**, 179–195.

Li, M., Kambhamettu, C., & Stone, M. 2005. Automatic contour tracking in ultrasound images. *Clinical Linguistics and Phonetics*, **19**(6–7), 545–554.

Zee, E. 1999. Change and variation in the syllable-initial and syllable-final consonants in Hong Kong Cantonese. *Journal of Cantonese Linguistics*, **27**, 120–167

The role of the tongue root in phonation of American English stops

Suzy Ahn (*New York University*)

Background. In American English, phonologically voiced consonants are often phonetically voiceless in utterance-initial position (Lisker & Abramson, 1964). Utterance-initial position is the context in which it is possible to test whether or not a language has stops with pre-voicing because 'active voicing' gestures by speakers are needed in this position (Beckman et al., 2013). Other than Westbury (1983), there is little articulatory evidence regarding utterance-initial voicing in American English. Westbury (1983) found that the tongue root is advanced in voiced consonants in utterance-initial positions, but he did not distinguish between phonated and unphonated voiced stops. The current study explores the question of what the phonetic target of voiced stops in English is and how the tongue root is employed to reach that phonetic target, comparing phonated voiced stops, unphonated voiced stops, and voiceless stops in utterance-initial position.

Hypothesis. One adjustment for initiating or maintaining phonation during the closure is enlarging the supraglottal cavity volume primarily via tongue root advancement (Westbury, 1983; Narayanan et al., 1995; Proctor et al., 2010). The same mechanism that is responsible for phonation during closure also facilitates short positive voice onset time (VOT) (Cho & Ladefoged, 1999). This study focuses on whether phonated voiced stops and unphonated voiced stops show the same tongue root position or not. If they are the same, it would suggest that speakers have the same phonetic target, i.e. short positive VOT, for both phonated and unphonated stops, but phonation can occur as a by-product of achieving that goal. If tongue positions are not the same, then it would suggest that speakers have phonation during closure as the phonetic target for phonated voiced stops.

Method. This study uses ultrasound imaging and acoustic measures to examine how tongue position corresponds to phonation in American English. Eight speakers of American English recorded voiced and voiceless stops in utterance-initial position at three places of articulation (labial, alveolar, and velar). For voiced stops, two different following vowels (high/low) were recorded. There were a total of 90 stimuli. Smoothing Spline (SS) ANOVA was used to compare the average contours between unphonated/phonated voiced and voiceless stops (Gu, 2002; Davidson, 2006).

Results. Acoustic results showed that there were 35 phonated stops out of 477 utterance-initial stops (7.3%). Ultrasound images showed that in utterance-initial position, there was a clear distinction between voiced stops and voiceless stops in the tongue root position for the alveolar and velar places of articulation. Labial stops do not participate in the pattern because they do not involve the tongue at all for the stop itself. Figures below demonstrate that both phonated (green curves) and unphonated (blue curves) voiced stops show more advanced tongue root than voiceless stops (orange curves) when the place of articulation is alveolar (Figure 1) or velar (Figure 2). Even without acoustic phonation during closure, the tongue root is advanced for voiced stops in comparison to voiceless stops for supraglottal cavity enlargement.

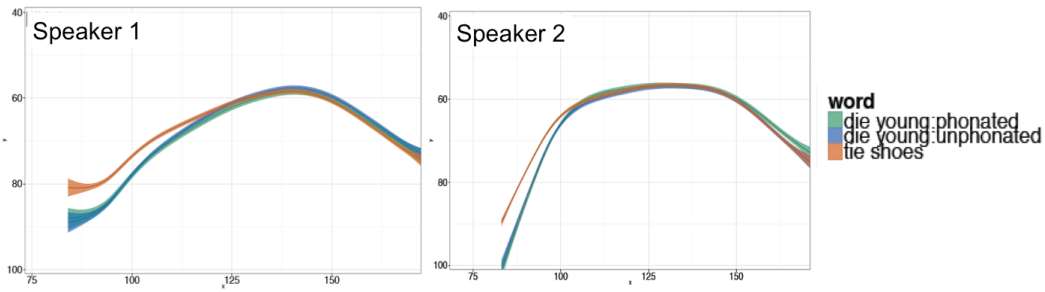


Figure 1. Phonated /d/ vs. unphonated /d/ vs. voiceless /t/ (SS ANOVA plots of two speakers)

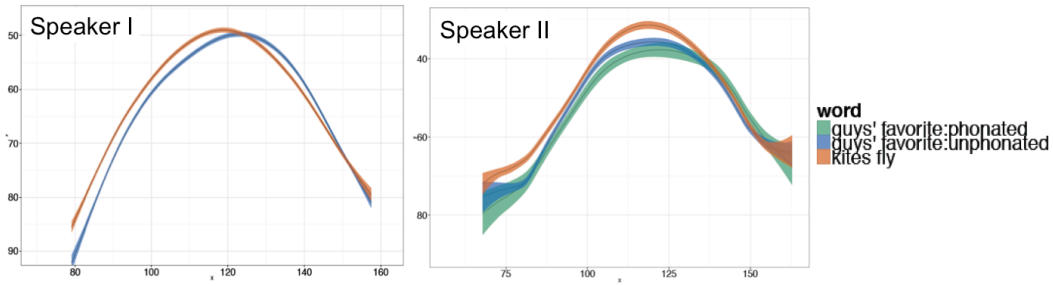


Figure 2. Phonated /g/ vs. unphonated /g/ vs. voiceless /k/ (SS ANOVA plots of two speakers, these are two different speakers from the speakers of *Figure 1*)

Discussion. These results are consistent with speakers having a short positive VOT as the target for both phonated and unphonated stops in utterance-initial position, but other articulatory adjustments are responsible for the presence or absence of phonation. One possible source of phonation may be hyper-articulation (Baese-Berk & Goldrick, 2009). (cf: hypercorrection in German: Jessen & Ringen, 2002)

Future Research (Pilot Study). The results found in English can be compared to other languages with different laryngeal feature systems, such as Spanish (a language with pre-voicing), German (a language similar to English), Thai or Hindi (a language with voiced/voiceless unaspirated/voiceless aspirated distinction), and Korean (a language without phonological voicing). A pilot study on Spanish showed that the tongue root is advanced in phonated voiced stops compared to (unaspirated) voiceless stops. English unphonated voiced stops are phonetically similar to Spanish unaspirated voiceless stops, but the tongue position is different in these two languages when they're both compared to the phonated voiced stop in their respective language. The difference is that in English, phonated and unphonated voiced stops are the same phoneme, whereas in Spanish, phonated voiced stops and unaspiraetd voiceless stops are different phoneme. This result indicates that the difference in tongue root position reflects the phonological laryngeal contrasts of English and Spanish, and phonation during closure in English is just accidental or entirely due to some other articulatory adjustment. A pilot study on Korean showed that the tongue root is advanced in tense stops, which have a shortest positive VOT, compared to lenis or aspirated stops, which have a longer VOT. These results confirm that tongue root advancement facilitates short positive VOT as well as phonation during closure. In this regard, German is expected to show the similar pattern to English, and Thai or Hindi are expected to show more tongue root advancement in voiced stops, followed by voiceless unaspirated stops, and then voiceless aspirated stops.

Reference:

Baese-Berk, Melissa & Matthew Goldrick (2009). Mechanisms of interaction in speech production. *Language and cognitive processes, 24*(4), 527-554.

Beckman, Jill, Michael Jessen & Catherine Ringen (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics, 49*(02), 259-284.

Cho, Taehong & Peter Ladefoged (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of phonetics, 27*(2), 207-229.

Davidson, Lisa (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variancea). *The Journal of the Acoustical Society of America, 120*(1), 407-415.

Gu, Chong (2002). *Smoothing Spline ANOVA Models*: Springer Science & Business Media.

Jessen, Michael & Catherine Ringen (2002). Laryngeal features in German. *Phonology, 19*(02), 189-218.

Lisker, Leigh & Arthur S Abramson (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*(3), 384-422.

Narayanan, Shrikanth S, Abeer A Alwan & Katherine Haker (1995). An articulatory study of fricative consonants using magnetic resonance imaging. *The Journal of the Acoustical Society of America, 98*(3), 1325-1347.

Proctor, Michael I, Christine H Shadle & Khalil Iskarous (2010). Pharyngeal articulation in the production of voiced and voiceless fricativesa). *The Journal of the Acoustical Society of America, 127*(3), 1507-1518.

Westbury, John R (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *The Journal of the Acoustical Society of America, 73*(4), 1322-1336.

Bolstering phonological fieldwork with ultrasound: lenition and approximants in Iwaidja

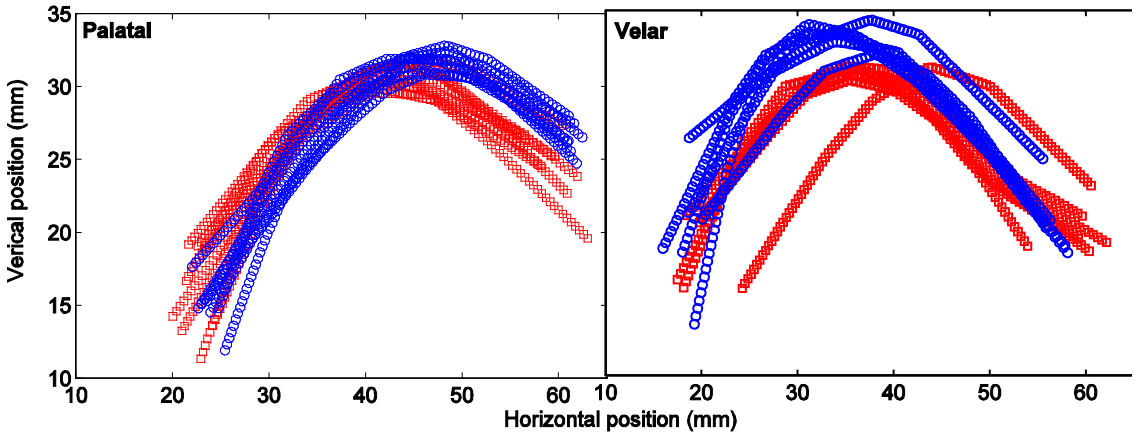
Robert Mailhammer[1], Mark Harvey[2], Tonya Agostini[1], Jason A. Shaw[1]

[1]Western Sydney University, [2]Newcastle University

Australian languages often have labial, palatal, and retroflex approximants. In addition, Iwaidja, an Australian language spoken in North-Western Arnhem Land, has a velar phoneme that has been analysed variably as either an approximant /ɯ/ (Evans 2009: 160) or a fricative /ɣ/ (Evans 2000: 99). This phoneme has a limited distribution, occurring only between [+continuant] segments. Across Australian languages, velar approximants commonly surface as an allophone of the velar stop in intervocalic position, where stops, particularly velar and labial stops, tend to undergo lenition. To ascertain the phonetic nature of the velar approximant in Iwaidja, in particular its status as an approximant (c.f. fricative) and its relation to lenited stops, we conducted the first instrumental phonetic investigation of Iwaidja, acquiring both acoustic and ultrasound data.

Ultrasound images and synchronized audio were collected in a field setting on Croker Island in the Northern Territory, Australia. Four speakers (1 female) participated in the study. Materials were designed to elicit the velar consonants [g, ɯ/ɣ], and also, as a comparison, the palatal stop-approximant contrast /ɟ, j/. Target words containing /g, ɯ/ɣ, ɟ, j/ in intervocalic position were elicited using objects pictured on a computer monitor. Ultrasound and audio data were recorded while participants named the pictures in a standardised carrier phrase. Ultrasound recordings were made with a GE 8C-RS ultrasound probe held at a 90 degree angle to the jaw in the mid-sagittal plane with a lightweight probe holder (Derrick et al., 2015). The probe was connected to a GE Logiq-E (version 11) ultrasound machine. Video output from the ultrasound machine went through an Epiphan VGA2USB Pro frame grabber to a laptop computer, which used FFMPEG running an X.264 encoder to synchronize video captured at 60Hz with audio from a Sennheiser MKH 416 microphone.

Preliminary analysis (see figure) indicates a clear distinction between articulation of consonants previously analysed as stops (blue circles) and as approximants (red squares) at both palatal (left panel) and velar (right panel) places of articulation. The figure compares edgetracks (Li et al. 2005) of 6-8 tokens per contrast in the same […a_a…] context. The origin of the plot is the posterior portion of the tongue. The stop [ɟ] (blue circles, left panel) differs from the approximant [j] (red squares, left panel) in being more front and slightly higher. The right panel shows the stop-approximant contrast at the velar place of articulation. Although the velar series is more variable than the palatal series, the velar stop is, on average, higher (~2mm) than the velar approximant. Acoustic data provides clear evidence of closure for palatal stops but not for velar stops. The height of the tongue for /ɯ ~ ɣ/ is similar to the vowel /u/ in our data. Although more analysis is required, preliminary results suggest that the velar contrast, which has been analysed as /g/ vs /ɯ/ or /ɣ/, is more accurately characterized as a contrast between /g/, which lenites to [ɯ], and a vowel /a/.



Selected References: Derrick, D., C. Best, R. Fiasson. (2015) Non-mettalic ultrasound probe holder for co-collection and co-registration with EMA. *Proceedings of ICPHS*; **Evans, N.** (2000). Iwaidjan, a very un-Australian language family. *Linguistic Typology*, *4*(2), 91-142.; **Evans, N.** (2009). Doubled up all over again: borrowing, sound change and reduplication in Iwaidjan. *Morphology* 19, 159-176; **Li, M., Kambhamettu, C., & Stone, M.** (2005). Automatic contour tracking in ultrasound images. *Clinical linguistics & phonetics*, *19*(6-7), 545-554.

Differences in the Timing of Front and Back Releases among Coronal Click Consonants

Amanda L. Miller
The Ohio State University

Clicks are multiply articulated consonants that have one constriction at the front of the mouth and another constriction at the back of the mouth. In coronal clicks, the front constrictions are produced by the tongue tip or blade contacting the hard palate, and the back constrictions are formed by the back of the tongue dorsum or uvula contacting the soft palate. Air is trapped in a lingual cavity between the two constrictions, and is rarefied by tongue body lowering and tongue dorsum retraction gestures, which differ among click types (Thomas-Vilakati 2010; Miller 2015a).

Differences in the timing of the coronal and dorsal releases in clicks have been deduced from acoustic properties of the bursts (Sands 1991; Johnson 1993). However, direct investigation of the timing of the two releases has not been previously undertaken. Ladefoged and Traill (1994) and Ladefoged and Maddieson (1996) note that it is necessary for the front release of a click to occur prior to the back release in order to rarefy the air and produce the "popping" sound that is characteristic of clicks. However, Stevens (1998) notes that while the front release in clicks generally occurs prior to the back release, some clicks have a more gradual front release with a distributed source. The current study investigates differences in the timing and the degree of opening of the coronal and dorsal releases in the four contrastive coronal clicks in the /i/ context in the Kx'a language Mangetti Dune !Xung using 114 fps ultrasound data collected using the CHAUSA method (Miller and Finch 2011). Results have implications for our understanding of two sound patterns in the Kx'a languages. The first is a C-V co-occurrence restriction, which is the basis for the complimentary distribution of [əi], which follows the alveolar and lateral clicks, and [i], which follows the dental and palatal clicks (Miller-Ockhuizen 2003). The second pattern is an innovative diachronic sound change from a palatal click in the proto language to a laminal alveolar click that occurs in the Northern branch of the Kx'a language family (Sands 2010; Miller and Holliday 2014).

The experiment presented here tests two hypotheses:

H1: Alveolar and lateral click types that retract and lower [i] to [əi] involve abrupt coronal releases with a large degree of opening; while the dental click type that co-occurs freely with [i] involves a more gradual front release that overlaps temporally with the back release.

H2: The palatal click type that occurs in [i] contexts has an abrupt release with a narrow opening resulting in secondary frication, which differs from the abrupt unfricated variant of the palatal click type with a wide opening that occurs preceding [ɑ].

The height of the tongue front and back at three time points, measured at 8.77 ms intervals over a 27 ms release phase that covers both the coronal and dorsal releases was measured from ultrasound tongue traces. The duration of different temporal phases of the click releases were also analyzed from acoustic data. Ultrasound and acoustic results support H1 by showing that the alveolar and lateral clicks, which co-occur with [əi], have more abrupt coronal releases that quickly change from a complete constriction to a wide aperture. The dental click, which occurs with [i], displays frication of the dental release that occurs due to a gradual front release with a very narrow aperture.

In keeping with H2, the results show that in the palatal click that co-occurs with [i], the front release barely opens to allow rarefaction, and then quickly returns to a more closed constriction resulting in secondary palatal frication. Thus, both clicks that co-occur with [i]

have more narrow front openings that overlap with the dorsal release. Conversely, the alveolar and lateral clicks that co-occur with [əi], have abrupt releases, leaving only the back constriction to overlap temporally with the following vowel.

The existence of the fricated palatal click variant is of great interest, as it provides evidence that there are two allophones of the palatal click type. The allophone of the palatal click with secondary palatal frication occurs in front vowel contexts (similar to other types of palatalization), while the abrupt variant of the palatal click occurs in back vowel contexts. Conversely, the dental click type is fricated in all contexts.

Differences in the timing of the front and back releases in clicks have implications for our understanding of how the lingual airstream mechanism works. The results also suggest a path for the development of the synchronic C-V co-occurrence restriction involving clicks and the high front vowel [i], as well as a possible path for sound change from a palatal click type to a fricated alveolar click type in the Kx'a language Ekoka !Xung (Miller 2015b) that is described by Miller and Holliday (2014).

References

Johnson, K. (1993). Acoustic and auditory analyses of Xhosa clicks and pulmonics. *UCLA working papers in phonetics 83, 33-45.*

Ladefoged, P. & Maddieson, I. (1996). *Sounds of the world's languages.* Cambridge, MA: Blackwell.

Ladefoged, P. & Traill, A. (1994). Clicks and their accompaniments. Journal of Phonetics, 22, 33-64.

Miller, A. (2015a). Posterior Lingual Gestures and Tongue Shape in Mangetti Dune !Xung Clicks. MS. The Ohio State University.

Miller, A. (2015b). Timing of the Two Release Gestures in Coronal Click Consonants. MS. The Ohio State University.

Miller, A. and Holliday, J. J. (2014). Contrastive apical post-alveolar and laminal alveolar click types in Ekoka !Xung. *Journal of the Acoustical Society of America* 135, 4, pp. 2351-2352.

Miller-Ockhuizen, A. (2003). *The Phonetics and phonology of gutturals: A case study from Ju|'hoansi*, In Horn, L. (Ed.), Outstanding Dissertations in Linguistics Series, New York: Routledge.

Sands, B. (2010). Juu subgroups based on phonological patterns. In Brenzinger, M. & König, C. (Eds.), Khoisan Languages and Linguistics. Proceedings of the 1st International Symposium January 4-8, 2003. Riezlern / Kleinwalsertal. Köln: Rüdiger Köppe Verlag.

Sands, B. (1991). *Evidence for click features: acoustic characteristics of Xhosa clicks. UCLA working papers in linguistics, 80, pp 6-37.*

Stevens, K. N. (1998). *Acoustic Phonetics*. Cambridge, MA.: MIT Press.

Thomas-Vilakati, K. (2010). *Coproduction and coarticulation in IsiZulu clicks*. University of California Publications in Linguistics. Volume 144. Berkeley and LosAngeles, CA. University of California Press.

Acoustic and Articulatory Speech Reaction Times with Tongue Ultrasound: What Moves First?

Pertti Palo, Sonja Schaeffler and James M. Scobbie

Clinical Audiology, Speech and Language (CASL) Research Centre, Queen Margaret University

1 Introduction

We study the effect that phonetic onset has on acoustic and articulatory reaction times. An acoustic study by Rastle et al. (2005) shows that the place and manner of the first consonant in a target affects acoustic RT. An articulatory study by Kawamoto et al. (2008) shows that the same effect is not present in articulatory reaction time of the lips. We have shown in a pilot study with one participant (Palo et al., 2015), that in a replication with Tongue Ultrasound Imaging (UTI), the same acoustic effect is present, but no such effect is apparent in the articulatory reaction time.

In this study we explore inter-individual variation with analysis of further participants. We also seek to identify the articulatory structures that move first in each context and answer the question whether this is constant across individuals or not.

2 Materials and methods

Since the phonetic materials, and recording and segmentation methods of this study are mostly the same as those we used in a previous study (Palo et al., 2015), we will provide only a short overview here. Three native Scottish English speakers (one male and two females) participated in this study. We carried out a partial replication of the Rastle et al. delayed naming experiment Rastle et al. (2005) with the following major changes: Instead of using phonetically transcribed syllables as stimuli, we used lexical monosyllabic words. The use of lexical words makes it possible to have phonetically naive participants in the experiment. In addition, we wanted to test if words with a vowel onset pattern in a systematic way with those with a consonant onset. Thus, the words were of /CCCVC/, /CCVC/, /CVC/, and /VC/ type.

The target words used in the original study were: *at*, *eat*, *ought*, *back*, *beat*, *bought*, *DAT*, *deep*, *dot*, *fat*, *feet*, *fought*, *gap*, *geek*, *got*, *hat*, *heat*, *hot*, *cat*, *keep*, *caught*, *lack*, *leap*, *lot*, *map*, *meet*, *mock*, *Nat*, *neat*, *not*, *pack*, *Pete*, *pop*, *rat*, *reap*, *rock*, *sat*, *seat*, *sought*, *shack*, *sheet*, *shop*, *tap*, *teak*, *talk*, *whack*, *wheat*, and *what*. For this study we added the following words with complex onsets: *black*, *drat*, *flat*, *Greek*, *crap*, *prat*, *shriek*, *steep*, *treat*, and *street*.

The experiment was run with synchronised ultrasound and sound recording controlled with Articulate Assistant Advanced (AAA) software Articulate Instruments Ltd (2012) which was also used for the manual segmentation of ultrasound videos. The participant was fitted with a headset to ensure stabilisation of the ultrasound probe Articulate Instruments Ltd (2008). Ultrasound recordings were obtained at a frame rates of ∼83 (for the first session with the male participant) and ∼121 (for all subsequent sessions) frames per second with a high speed Ultrasonix system. Sound was recorded with a small Audio Technica AT803b microphone, which was attached to the ultrasound headset. The audio data was sampled at 22,050 Hz.

Each trial consisted of the following sequence: (1) The participant read the next target word from a large font print out. (2) When the participant felt that they were ready to speak the word, they activated the sound and ultrasound recording by pressing a button on a keyboard. (4) After a random delay which varied between 1200 ms and 1800 ms, the computer produced a go-signal – a 50 ms long 1000 Hz pure tone.

The acoustic recordings were segmented with Praat Boersma and Weenink (2010) and the ultrasound recordings were segmented with AAA Articulate Instruments Ltd (2012) as in our previous study.

3 Pixel difference

Regular Pixel Difference (PD) refers simply to the Euclidean distance between two consecutive ultrasound frames. It is based on work by McMillan and Corley (2010), and Drake et al. (2013a,b). Our version of the algorithm is explained in detail by Palo et al. (2014).

Instead of using the usual interpolated ultrasound images in the calculations, we use raw uninterpolated images (Figure 1). The fan image of the ordinary ultrasound data is produced by interpolation between the actual raw data points produced by the ultrasound system. The raw data points are distributed along radial scanlines with the number of scanlines and the number of data points imaged along each scanline depending on the setup of the ultrasound system. In this study we obtained raw data with 63 scanlines covering an angle of about 135 degrees and with 256 pixels along each scanline.

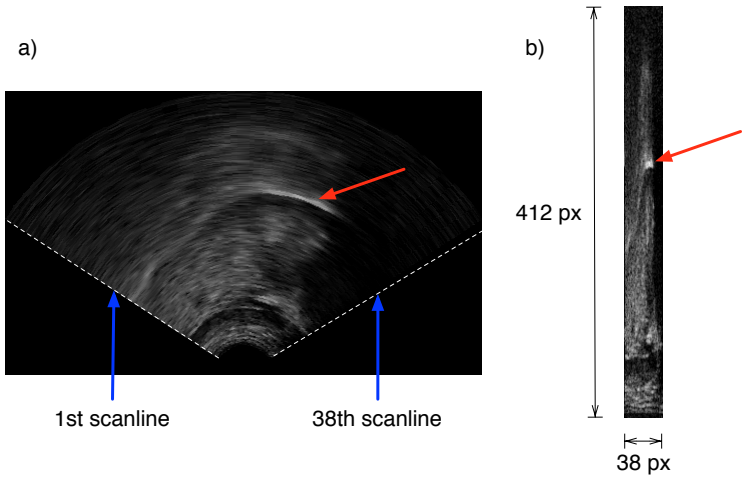


Figure 1: The difference between interpolated and raw ultrasound frames: a) An interpolated ultrasound frame. b) Raw (uninterpolated) version of the same ultrasound frame as in a). The speaker is facing right. Red arrow points to the upper surface of the tip of the tongue.

In addition to the overall frame-to-frame PD and more importantly for the current study, we also calculate the PD for individual scanlines as a function of time. This makes it possible to identify the tongue regions that initiate movement in a given token. Figure 2 shows sample analysis results. The lighter band in the middle panels around scanlines 53-63 is caused by the mandible, which is visible in ultrasound only as a practically black area with a black shadow extending behind it. This means that there is less change to be seen in most frame pairs in these scanlines than there is in scanlines which only image the tongue and its internal tissues.

As can be seen for the token on left ('caught'), the tongue starts moving more or less as a whole. In contrast the token on the right ('sheet') shows an early movement in the pharyngeal region before activation spreads to the rest of the tongue. This interpretation should be taken with (at least) one caveat: The PD does not measure tongue contour movement. This means that a part of the tongue contour might be the first to move even if the scanline based PD shows

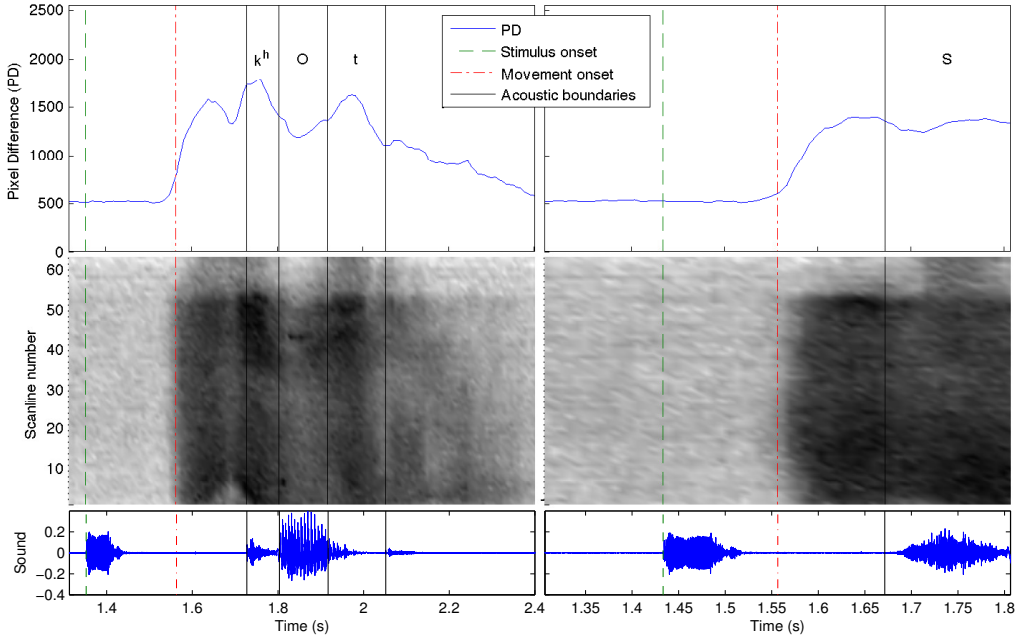


Figure 2: Two examples of regular PD and scanline based PD. The left column has a repetition word 'caught' ([kɔːt]) and the right column has the beginning of the word 'sheet' ([ʃiːt]). The panels are from top to bottom: Regular PD with annotations from acoustic segmentation, scanline based PD with the back most scanline at the bottom and the front most on top with darker shading corresponding to more change, and the acoustic waveform.

activation everywhere. This is because the PD as such measures change from frame to frame (whether on scanlines or on the whole frame). More detailed analysis will be available at the time of the conference.

References

Articulate Instruments Ltd (2008). *Ultrasound Stabilisation Headset Users Manual: Revision 1.4*. Edinburgh, UK: Articulate Instruments Ltd.

Articulate Instruments Ltd (2012). *Articulate Assistant Advanced User Guide: Version 2.14*. Edinburgh, UK: Articulate Instruments Ltd.

Boersma, P. and Weenink, D. (2010). Praat: doing phonetics by computer [computer program]. Version 5.1.44, retrieved 4 October 2010 from `http://www.praat.org/`.

Drake, E., Schaeffler, S., and Corley, M. (2013a). Articulatory evidence for the involvement of the speech production system in the generation of predictions during comprehension. In *Architectures and Mechanisms for Language Processing (AMLaP)*, Marseille.

Drake, E., Schaeffler, S., and Corley, M. (2013b). Does prediction in comprehension involve articulation? evidence from speech imaging. In *11th Symposium of Psycholinguistics (SCOPE)*, Tenerife.

Kawamoto, A. H., Liu, Q., Mura, K., and Sanchez, A. (2008). Articulatory preparation in the delayed naming task. *Journal of Memory and Language*, 58(2):347 – 365.

McMillan, C. T. and Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition*, 117(3):243 – 260.

Palo, P., Schaeffler, S., and Scobbie, J. M. (2014). Pre-speech tongue movements recorded with ultrasound. In *10th International Seminar on Speech Production (ISSP 2014)*, pages 304 – 307.

Palo, P., Schaeffler, S., and Scobbie, J. M. (2015). Effect of phonetic onset on acoustic and articulatory speech reaction times studied with tongue ultrasound. In *Proceedings of ICPhS 2015*, Glasgow, UK.

Rastle, K., Harrington, J. M., Croot, K. P., and Coltheart, M. (2005). Characterizing the motor execution stage of speech production: Consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5):1083 – 1095.

Keynote 2: Wednesday, December 9, 1:30-2:30pm

Patrick Wong
The Chinese University of Hong Kong

Neurophysiology of Speech Perception: Plasticity and Stages of Processing

Even after years of learning, many adults still have difficulty mastering a foreign language. While the learning of certain aspects of foreign languages, such as vocabulary, can be acquired with nearly native-like proficiency, foreign phoneme and phonological grammar learning can be especially challenging. Most interestingly, adults differ to a large extent in how successfully they learn. In this presentation, I will discuss the potential neural foundations of such individual differences in speech learning, including the associated cognitive, perceptual, neurophysiological, neuroanatomical, and neurogenetic factors, paying particular attention to the contribution of stages of processing along the auditory neural pathway. I will then describe a series of experiments that demonstrate that re-designing a learner's training protocol based on biobehavioral markers can sometimes optimize learning.

/r/-allophony and gemination: an ultrasound study of gestural blending in Dutch

Patrycja Strycharczuk[1], Koen Sebregts[2]
[1]CASL, Queen Margaret University, [2]Utrecht University

Standard Dutch increasingly displays an /r/ allophony pattern in which coda /r/ (e.g. *paar* 'couple') is realised as a post-alveolar approximant (bunched or retroflex), whereas onset /r/ (e.g. *raden* 'guesses') is typically a uvular fricative or trill (Scobbie and Sebregts 2010). In this paper, we investigate the spatial and temporal characteristics of coarticulation between these distinct allophones in a "fake geminate" context (*paar raden*). Fake geminates tend to undergo gradient degemination in Dutch (Martens and Quené 1994). However, while the /r#r/ sequence consists of phonemically identical consonants, they are phonetically strongly disparate. This invites the question of whether degemination also applies here, and in case it does, what it entails in gestural terms.

We present articulatory data from 4 speakers of Standard Dutch (3 females), collected with a high-speed ultrasound system (121 fps). The test materials included /r/ in canonical onset, canonical coda and fake geminate contexts, in a controlled prosodic and segmental environment (10 tokens per context per speaker). The ultrasound data were analysed using two methods: i) dynamic analysis of principal components of pixel intensity data in the ultrasound image (TRACTUS, Carignan 2014), and ii) SS-ANOVA (Davidson 2006) comparison of tongue contours at the point of maximal constriction for the /r/ and at the acoustic onset of the vowel. We used the principal components (PCs) obtained using TRACTUS in a Linear Discriminant Analysis trained to distinguish /aː#rV/ (*pa raden*) from /aːr#C/ (*paar baden*). We then used the algorithm to classify /r/ tokens in the fake geminate context, /aːr#r/, (*paar raden*). The average discriminant values for an example speaker, DF2, are plotted in Figure 1.

Figure 1

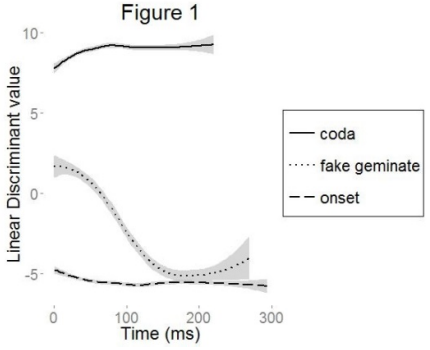
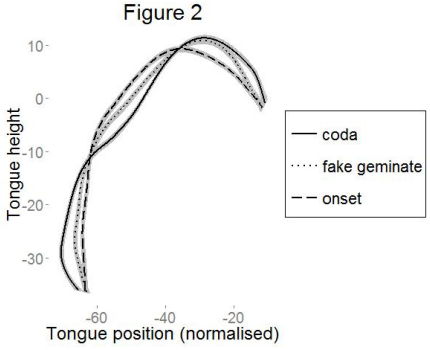


Figure 2



For most of the /aːr/ duration, the fake geminate context shows values that are in between the two baselines, suggesting an intermediate articulation between coda and onset /r/. This is confirmed by results of SS-ANOVA at the /r/-constriction: there is a simultaneous bunching gesture (as in canonical codas) and dorsal raising (as in canonical onsets) in *paar raden*, although both gestures are spatially reduced compared to those in non-geminate onsets and codas (Figure 2). In temporal terms, however, the fake geminate context shows no increase in duration compared to singleton onset /r/. In other words, the effect of degemination is strongest in the temporal domain. This situation is reminiscent of that of /l#l/ fake geminates in English (e.g. *peel lemurs*, Scobbie and Pouplier 2010), although these

show incomplete overlap and less temporal reduction. The Dutch facts can be captured in Articulatory Phonology (AP) as a blending of two gestures that overlap completely in time. We discuss such an interpretation in the context of the restrictive view AP takes towards allophony (two allophones are considered to consist of the same gestures, with possible differences in magnitude and timing), which is problematised by the Dutch allophonic [ʁ]~[ɹ] pattern.

References

Carignan, C. (2014). TRACTUS (Temporally Resolved Articulatory Configuration Tracking of Ultrasound) software suite, http://phon.chass.ncsu.edu/tractus/

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing/spline analysis of variance. *The Journal of the Acoustical Society of America 120*, 407-415.

Martens, L. & H. Quené (1994). Degemination of Dutch fricatives in three different speech rates. In: R. Bok-Bennema and C. Cremers (Eds.), *Linguistics in the Netherlands 1994* (pp. 119-126). Amsterdam: John Benjamins.

Scobbie, J.M. and M. Pouplier (2010). The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction in word-final English /l/. *Journal of Phonetics 38*(2), 240-259.

Scobbie, J.M. and K. Sebregts (2010). Acoustic, articulatory and phonological perspectives on allophonic variation of /r/ in Dutch. In: R. Folli, & C. Ulbrich (Eds.), *Interfaces in Linguistics: New Research Perspectives*. Oxford: Oxford University Press.

Allophonic variation: An articulatory perspective

Alessandro Vietti, Lorenzo Spreafico, Vincenzo Galatà

Free University of Bozen-Bolzano

1. Introduction

In this paper, we explore the issue of allophonic variation via a quantitative and qualitative analysis of /r/ in Tyrolean -- a South Bavarian Dialect --. The allophony of /r/ in this High German language is a challenging problem and only a few attempts have been done to solve it, usually basing on acoustic and articulatory descriptions of all attested /r/-variants or on their contextual distribution. Interestingly, most previous researches have highlighted a high degree of intraspeaker variation in the uvular realizations of the rhotics [1]. Hence, here we provide novel UTI data on Tyrolean to discuss both the "phonological allophony" -- namely the variation "predictably conditioned by categorically distinct phonological contexts" --, and the "phonetic allophony" -- namely the "cases of predictable contextual differences which exist but which are not thought to be represented by changing the internal phonological content of segments" [2] --.

2. Methodology

For the analysis, we employed acoustic and ultrasonic data synchronized using the Articulate Assistant Advanced (AAA) software package [3]. Tongue profiles were captured by means of an Ultrasonix SonicTablet ultrasound imaging system. Tongue contours were tracked using the Ultrasonix C9-5/10 transducer operating at 5MHz. Ultrasound recordings were collected at a rate of about 90Hz with a field of view of about 120°. Acoustic data were recorded by means of a Sennheiser ME2 microphone connected to a B1 Marantz PMD660. Audio was sampled at 22050Hz 16-bit mono.

The stimuli included **80 real Tyrolean words**, eliciting /r/ in all possible syllable contexts and positions (onset vs. coda, simple vs. complex, initial vs. medial vs. final) according to an in-depth scrutiny of all available dictionaries of contemporary Tyrolean. In compiling the word list, surrounding vowels (V) were restricted to /a, i, o/; surrounding consonants (C) for /r/ in syllable onset (CRV) and coda (VRC) position were restricted to /t, d, k, g/. For /r/ in coda position words with /r/ + nasal or liquid were also included [4].

Five native Tyrolean speakers with no reported speech disorders were recorded. Participants were aged between 25 and 35 and were born and living in the area of Meran. All subjects had command of Tyrolean as well as of Standard German and Standard Italian at native-like level.

3. Analysis

The preliminary acoustic-auditory labelling process identifies four possible uvular /r/-variants (trill, tap, fricative and approximant) plus a vocalized variant. The variants are not equally distributed in the sample and do not strictly correlate with the phonetic contexts. However, the following trends emerge: the fricative is the default choice; trills and taps are more likely to occur in onset contexts, the process of r-vocalization is restricted to the coda position. Trends are computed using a multivariate approach to the analysis of data [5].

Fitted splines taken from the acoustic midpoint of each labelled /r/-variant were exported to the AAA's workspace in order to calculate the smoothed tongue contour for each variant in each speaker. The analysis was run in R according to [6]. The comparison of /r/-variants profiles irrespective of the phonetic contexts they were in shows that notwithstanding marked allophonic variation in the acoustics, the articulatory patterns are relatively stable (fig. 1).

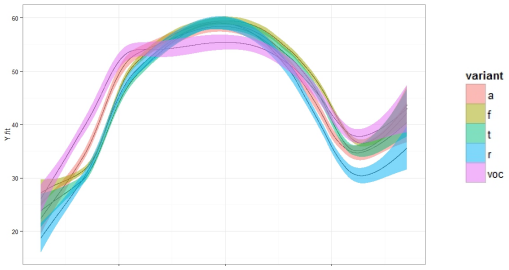


Figure 1: Smoothing splines results for SP1's /r/-variants (colour legend on the left in the following order: a = approximant, f = fricative, t = tap, r = trill, voc = vocalization).

The investigation of extracted tongue profiles shows an overall similarity in tongue shape and position regardless of coarticulatory effects. In particular, the following parameters seem to be contributing to the overall /r/ tongue profiles hence to the allophony.

(1) The degree of dorsal constriction (t > f > a > v, similarly to what proposed in [8, 9] in regard to the articulatory unity of German /r/);

(2) The peculiar combination of root retraction, tongue blade lowering and tongue dorsum bunching.

Collected data will be used to discuss the phonological vs. phonetic allophony of Tyrolean, and to address the more general question of allophony from the standpoint of articulatory phonetics.

[1] Spreafico, L., Vietti, A. 2013. On rhotics in a bilingual community: A preliminary UTI research. In: Spreafico, L., Vietti, A. (eds.), *Rhotics. New data and perspectives*. BU Press, 57-77.

[2] Scobbie, J., Sebregts, K. 2011. Acoustic, articulatory and phonological perspectives on rhoticity and /r/ in Dutch. In: Folli, R., Ulbrich, C. (eds.), *Interfaces in linguistics: new research perspectives*. OUP, 257-277.

[3] Articulate Instruments Ltd 2014. *Articulate Assistant Advanced User Guide: Version 2.15*. Edinburgh, UK: Articulate Instruments Ltd.

[4] Vietti A., Spreafico, L. Galatà, V. 2015. A*n ultrasound study on the phonetic allophony of tyrolean /r/. ICPhS 2015.*

[5] Vietti A., Spreafico, L. (in press). *Lo strano caso di /R/ a Bolzano: problemi di interfaccia.*. In: Claudio Iacobini (eds.), *Livelli di analisi e interfaccia*. Roma, Bulzoni.

[6] Davidson, L. 2006. Comparing tongue shapes from ultrasound imagining using smoothing spline analysis of variance. *JASA* 120(1), 407-415.

[7] Wiese, R. 2000. *The Phonology of German*. Oxford: OUP.

[8] Schiller, N. 1998. The phonetic variation of German /r/. In Butt M., Fuhrhop N. (eds.) *Variation und Stabilität in der Wortstruktur.* Olms: 261-287.

[9] Klein, K., Schmitt, L. 1969. *Tirolischer Sprachatlas*. Tyrolia-Verlag.

Taps vs. Palatalized Taps in Japanese

Noriko Yamane & Phil Howson
University of British Columbia & University of Toronto

This paper examines the dynamic mid-sagittal lingual contrast between the plain and palatalized taps in Japanese. Japanese taps are basically same as English flap such as in 'ladder' (Vance 1997), but the kinematics of the movement hasn't been paid much attention. Although Japanese tap allows allophonic/sociophonetic variants such as apico-alveolar lateral [l̺], voiced alveolar lateral fricative [ɮ], Retroflex [ɽ], and apical trills [r] in adults (Magnuson 2010, Labrune 2012), the canonical Japanese taps are challenging even for native speakers of Japanese (e.g, Ueda 1996). Japanese taps are challenging for English speakers as well, although English taps also allow variants such as alveolar/postalveolar taps and down/up flaps (Derrick & Gick 2011). Japanese palatalized taps seem more challenging (Tsurutani 2004), which is likely related to the cross-linguistic rarity of palatalized tap (Hall 2000). This paper explores why these sounds are challenging from the viewpoint of articulatory kinematics, using ultrasound. Taps in Japanese have not been well research using articulatory methods; therefore, the primary goal of this paper is to reveal the articulatory dynamics of taps in Japanese. Palatalized taps are also typologically rare, as are any palatalized rhotics.

Six native speakers of Japanese participated in ultrasound experiment, and produced nonsense words containing /ɾ/ and /ɾʲ/ in a carrier sentence. The mid-sagittal contours of the taps were compared in three intervocalic contexts: a_a, o_o, u_u. Static measures at the point of contact were compared as dynamic measures of the movements over time. For the static measure, images were extracted at the point of tongue tip contact, which was determined by a spectral occlusion in the spectrogram. The dynamic measures were taken from the spectral occlusion: 4 frames before the occlusion, on the spectral occlusion, and 5 frames after the occlusion, for a total of 10 images. Due to the frame rate of the ultrasound, images are approximately 33 ms apart. Results were compared in R (R Core Development Team 2015) using an SSANOVA (Davidson 2006).

The results indicate that /ɾʲ/ is more resistant to coarticulatory effects of adjacent vowels compared to /ɾ/. Both the apical gesture and the tongue body gesture were invariable regardless of vocalic environment. /ɾ/ was articulated with a very brief occlusion by tongue tip (Figure 1), while /ɾʲ/ was articulated with tongue tip raising followed by tongue body raising and fronting (Figure 2). However, unlike palatalized trills, there doesn't seem to be a coarticulatory conflict between the tongue dorsum and palatalization. This is largely because the tongue dorsum for /ɾ/, showed a high degree of coarticulatory variability with the surrounding vocalic environment, suggesting that there is no tongue dorsum gesture involved in taps, similar to Catalan (Recasens & Espinosa 2007). The resistance of the marked counterpart of the tap against conflicting vowel context is also similar to Catalan (Recasens and Pallarès 1999).

The results also suggest that an inconsistency between palatalization and rhotics cannot be related to the constraints on the dorsal gesture as Kavitskaya et al. (2009) suggest, because the dorsal gesture seems to be inert for the taps. Rather phonological contrast within liquids (e.g., Scobbie et al. 2013, Proctor 2009) should be considered.

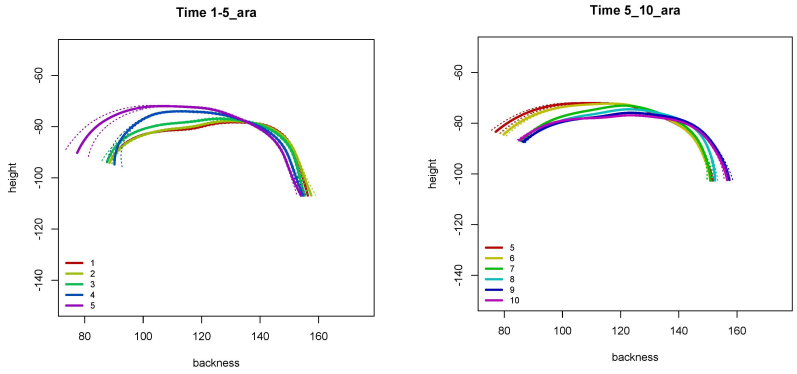


Figure 1. Left: closing gesture from /a1/ to tap. Right: opening gesture from tap to /a2/. SSANOVAs from 12 tokens for each time frame from one female speaker. Tongue tip is on the left side of the images.

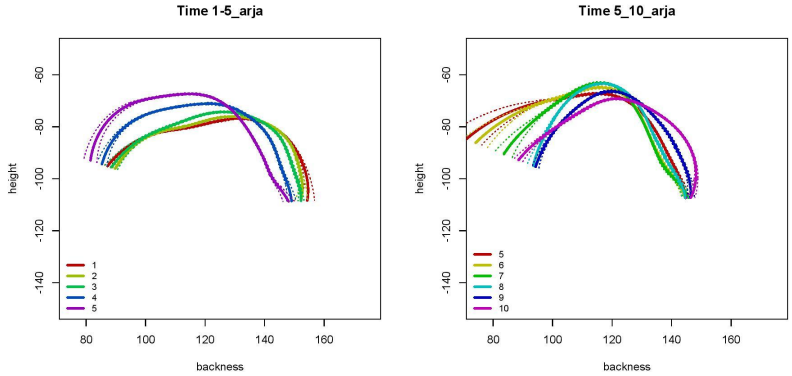


Figure 2. Left: closing gesture from /a1/ to palatalized tap. Right: opening gesture from palatalized tap to /a2/. SSANOVAs from 12 tokens for each time frame from one female speaker. Tongue tip is on the left side of the images.

Acknowledgements

This project is supported by a Flexible Learning Large Project Grant from the Teaching and Learning Enhancement Fund at the University of British Columbia.

References

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance). *The Journal of the Acoustical Society of America*, *120*(1), 407-415.

Derrick, D., & Gick, B. (2011). Individual variation in English flaps and taps: A case of categorical phonetics. *The Canadian Journal of Linguistics/La revue canadienne de linguistique*, *56*(3), 307-319.

Hall, T. A. (2000). Typological generalizations concerning secondary palatalization.Lingua 110, 1-25.

Labrune, L. (2014). The phonology of Japanese/r: a panchronic account.*Journal of East Asian Linguistics*, *23*(1), 1-25.

Proctor, M. 2009. *Gestural characterization of a phonological class: the liquids*. New Haven, CT: Unpublished Ph.D. dissertation. Yale University, New Haven.

Recasens, D., & Espinosa, A. (2007). Phonetic typology and positional allophones for alveolar rhotics in Catalan. *Phonetica*, *64*(1), 1-28.

Recasens, D., & Pallarès, M. D. (1999). A study of /r/ and /ɾ/ in the light of the DAC coarticulation model. *Journal of Phonetics*, *27*(2), 143-169.

Ueda, I (1996). Segmental acquisition and feature specification in Japanese. (eds.) B.Bernhardt, J. Gilbert and D. Ingram Proceedings of the UBC.

International Conference on Phonological Acquisition, 15-24. Somerville, MA.:Cascadilla Press.

Magnuson, T. (2010). A Look into the Plosive Characteristics of Japanese/r/and/d. *Canadian Acoustics*, *38*(3), 130-131.

Tsurutani, C. (2004). Acquisition of Yo-on (Japanese contracted sounds) in L1 and L2 phonology. *Second Language*, *3*, 27-47.

Kavitskaya, D., Iskarous, K., Noiray, A., & Proctor, M. (2009). Trills and palatalization: Consequences for sound change. *Proceedings of the formal approaches to slavic linguistics*, *17*, 97-110.

Scobbie, J. M., Punnoose, R., & Khattab, G. (2013). Articulating five liquids: A single speaker ultrasound study of Malayalam.

Vance, T. J. (1997). *An introduction to Japanese phonology*. SUNY Press.

Russian palatalization, tongue-shape complexity measures, and shape-based segment classification

Kevin D. Roon[1,2], Katherine M. Dawson[1,2], Mark K. Tiede[2,1], D. H. Whalen[1,2,3]
[1]CUNY Graduate Center, [2]Haskins Laboratories, [3]Yale University

The present study will address two research goals by analyzing ultrasound images of utterances from Russian speakers. The first goal is to provide a better characterization of the articulation of palatalized vs. non-palatalized consonants than is currently available. The second is to test and extend the shape analyses developed by Dawson, Tiede, and Whalen (accepted). One set of CVC stimuli contains palatalized and non-palatalized consonants in word-initial and word-final positions. Another set contains all of the vowels of Russian.

The most extensive ultrasound study of Russian palatalized consonants is Proctor (2011), which reports head-corrected ultrasound data (Whalen et al., 2005) for the palatalized and non-palatalized liquids /r/ and /l/, as well as /d/, in three vowel contexts (/e, a, u/). The present study differs from the Proctor (2011) study in two ways. First, Proctor (2011) was primarily concerned with characterizing liquids, whereas the present study will be primarily concerned with characterizing palatalization. Second, the present study will investigate palatalization in consonants with a greater number of primary oral articulators, manners, and word positions than Proctor (2011).

Dawson et al. (accepted) compared new and previously used methods for quantifying the complexity of midsagittal tongue shapes obtained with ultrasound. In that study, the first coefficient of a Fourier shape analysis similar to that of Liljencrants (1971) was used to successfully classify the consonants in *aCa* utterances and vowels in *bVb* utterances produced by English speakers based on the shape alone, that is, without any information about the position of the tongue in the vocal tract. The present study will test and extend the analyses from Dawson et al. (accepted) in two ways. First, we will compare the complexity and classification results from Russian vowels and non-palatalized consonants with the results for English. Second, we will investigate what the effects of both palatalization and word position (and the combination of the two) are on these complexity and classification measurements.

References

Dawson, K. M., Tiede, M. K., & Whalen, D. H. (accepted). Methods for quantifying tongue shape and complexity using ultrasound imaging. *Clinical Linguistics & Phonetics*.

Liljencrants, J. (1971). Fourier series description of the tongue profile. *Speech Transmission Laboratory-Quarterly Progress Status Reports, 12*(4), 9–18.

Proctor, M. (2011). Towards a gestural characterization of liquids: Evidence from Spanish and Russian. *Laboratory Phonology, 2*(2), 451–485.

Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E., & Hailey, D. S. (2005). The Haskins Optically Corrected Ultrasound System (HOCUS). *Journal of Speech, Language, and Hearing Research, 48*, 543–553.

Exploring the relationship between tongue shape complexity and coarticulatory resistance

D. H. Whalen[1,2,3], Kevin D. Roon[1,2], Katherine M. Dawson[1,2], Mark K. Tiede[2,1]
[1]CUNY Graduate Center, [2]Haskins Laboratories, [3]Yale University

Coarticulation, the influence of one segment on another, is extensive in speech, and is a major source of the great variability found in speech (e.g., Iskarous, et al., 2013; Öhman, 1967). Consonants have been found to allow or "resist" coarticulation to varying degrees (e.g., Fowler, 2005; Recasens, 1985). Correlates of coarticulatory resistance have been found in tongue position (Recasens & Espinosa, 2009) and jaw height (Recasens, 2012). Our aim in the present study is to see whether there is a relationship between tongue shape and resistance to coarticulation. To this end, we have collected data from one speaker of English (with three more planned) producing VCV nonsense strings. The Vs were symmetrical /ɑ/, /i/ or /u/. The Cs were one of the group /m p n t k r l s ʃ/. These were repeated 20 times in random order with optically corrected ultrasound imaging (HOCUS; Whalen, et al., 2005). Tongue shapes were measured with GetContours (Haskins Labs) and quantified via the measures described in Dawson et al. (submitted). The nine consonants will be ranked by the quantified measures of tongue shape and complexity, and that ranking will be compared with the ranking of coarticulatory resistance generated from the various articulatory and acoustic studies of that phenomenon.

Dawson, K. M., Tiede, M. K., & Whalen, D. H. (submitted). Methods for quantifying tongue shape and complexity using ultrasound imaging. *Clinical Linguistics and Phonetics*.

Fowler, C. A. (2005). Parsing coarticulated speech in perception: Effects of coarticulation resistance. *Journal of Phonetics, 33*, 199-213.

Iskarous, K., Mooshammer, C. M., Hoole, P., Recasens, D., Shadle, C. H., Saltzman, E., & Whalen, D. H. (2013). The Coarticulation/Invariance Scale: Mutual Information as a measure of coarticulation resistance, motor synergy, and articulatory invariance in speech. *Journal of the Acoustical Society of America, 134*, 1271-1282.

Öhman, S. E. G. (1967). Numerical model of coarticulation. *Journal of the Acoustical Society of America, 41*, 310-320.

Recasens, D. (1985). Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences. [Article]. *Language and Speech, 28*, 97-114.

Recasens, D. (2012). A study of jaw coarticulatory resistance and aggressiveness for Catalan consonants and vowels. *Journal of the Acoustical Society of America, 132*, 412-420.

Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *Journal of the Acoustical Society of America, 125*, 2288-2298.

Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E., & Hailey, D. S. (2005). HOCUS, the Haskins Optically-Corrected Ultrasound System. *Journal of Speech, Language, and Hearing Research, 48*, 543-553.

An investigation of lingual coarticulation resistance using ultrasound

Daniel Recasens & Clara Rodríguez
Universitat Autònoma de Barcelona & Institut d'Estudis Catalans, Barcelona, Spain

Introduction

This paper uses ultrasound data in order to explore the extent to which lingual coarticulatory resistance for front lingual consonants and vowels in VCV sequences increases with the place and manner of articulation requirements involved in their production. Coarticulatory resistance for a given consonant or vowel is a measure of its degree of articulatory variability as a function of phonetic context such that the less the target segment adapts to the articulatory configuration for the flanking segments, the more coarticulation resistant it may be assumed to be. In principle, ultrasound should be more appropriate than EPG and EMA for studying coarticulatory resistance since it allows us to measure phonetic contextual effects not only at the alveolar and palatal zones but at the velar zone and at the pharynx as well.

In the present investigation coarticulatory resistance will be evaluated for the Catalan consonants /t, d, n, l, s, ɾ, r, ʎ, ɲ, ʃ/ and vowels /i, e, a, o, u/ embedded in symmetrical VCV sequences. In present-day Catalan, those consonants may be characterized as follows: /t, d/ are dentoalveolar and /d/ is realized as an approximant intervocalically ([ð]); among the alveolar consonants /n, l, s, ɾ, r/, /ɾ/ is a tap, /r/ is a trill and /l/ is clear rather than dark (as for the Catalan speakers who took part in the present study, F2 for /l/ amounts to 1400 Hz next to /i, e/ in the case of males and to 2500 Hz next to /i/ and 1700 Hz next to /e/ in the case of females); /ʃ/ is palatoalveolar and /ʎ, ɲ/ are alveolopalatal.

Within the framework of the degree of articulatory constraint (DAC) model of coarticulation and in line with kinematic data reported elsewhere (Recasens & Espinosa, 2009), we hypothesized that the degree of coarticulatory resistance for the phonetic sounds under investigation ought to conform to specific trends. On the one hand, palatal consonants and palatal vowels were expected to be most resistant since their production involves the entire tongue body. On the other hand, coarticulatory resistance for dentoalveolar consonants should depend on manner of articulation and thus, be highest for /s/ and the trill /r/, lowest for the approximant [ð], and intermediate for /t, n, ɾ/ and clear /l/. As for vowels, differences in tongue constriction location and lip rounding should render /a/ less variable than /o, u/. In sum, our initial hypothesis was that coarticulatory resistance ought to decrease in the progression /ʎ, ɲ, ʃ/ > /t, n, ɾ, l/ > /s, r/ > /d/ for consonants and /i, e/ > /a/ > /o, u/ for vowels.

Method

The speech materials, i.e., symmetrical VCV sequences with /t, d, n, l, s, ɾ, r, ʎ, ɲ, ʃ/ and /i, e, a, o, u/, were recorded by five native speakers of Catalan, three females and two males, wearing a stabilization headset. Tongue contours were tracked automatically and adjusted manually every 17.5 ms with the Articulate Assistant Advanced program. The resulting 83 data point splines were then exported as X-Y coordinates, converted from Cartesian into polar coordinates, and submitted to a smoothing SSANOVA computation procedure (Davidson 2006, Mielke, 2015). Based on EPG data on constriction location for specific Catalan consonants (Recasens, 2014) and on vocal tract morphology data available in the literature (Fitch & Giedd, 1999), the splines in question were subdivided into four portions which correspond to the alveolar, palatal, velar and pharyngeal articulatory zones (see Figure 1). As revealed by the graph, the articulatory zones differed in size in the progression pharygeal > velar, palatal > alveolar for all speakers.

Coarticulatory resistance was measured at each articulatory zone for consonants at C midpoint using the mean splines across tokens for the five contextual vowels /i, e, a, o, u/, and for vowels at the V1 and V2 midpoints using the mean splines across tokens for the ten contextual consonants /t, d, n, l, s, ɾ, r, ʎ, ɲ, ʃ/. It was taken to equal the area of the polygon embracing all contextual splines as determined by the maximal and minimal Y values at all points along the X axis (Figure 1 shows the polygon for /l/ at the palatal zone for exemplification). In all cases, the smaller the area of the polygon, the higher the degree of coarticulatory resistance. In order to draw interspeaker comparisons the area values of the polygons computed with Gauss' formula were submitted to a normalization procedure separately at each articulatory zone by subtracting the mean area value across all consonants or vowels from the area value for each individual consonant or vowel and dividing the outcome by the standard deviation of the mean.

The resulting normalized area values were submitted to an ANOVA analysis with 'consonant' or 'vowel' and 'zone' as fixed factors and 'subject' as a random factor. The statistical results will be interpreted with reference to the 'consonant' or 'vowel' main effect and the 'consonant'/'vowel' x 'zone' interaction but not to the 'zone' main effect since the normalization procedure happened to level out the differences in area size among the polygons located at differents zones (see above).

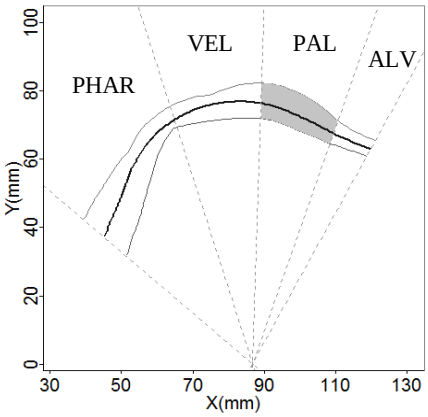


Figure 1. Subdivision of the lingual spline field for /l/ into the four articulatory zones ALV (alveolar), PAL (palatal), VEL (velar) and PHAR (pharyngeal). The spline field encompasses the splines for /ili, ele, ala, olo, ulu/. The polygon for the palatal zone is highlighted for exemplification.

Results

The statistical results for the consonant data yielded a main effect of 'consonant' ($F(9, 160)=80.39$, $p< 0.001$) and a 'consonant' x 'zone' interaction ($F(27, 160)=3.09$, $p< 0.001$). As shown in Figure 2, a Tukey post-hoc test revealed that the area size across zones varies in the progression /d/ ([ð]) > /l, ɾ, t, n/ > /s, r/ > /ʎ, ɲ, ʃ/ and simple effects tests that these consonant-dependent differences hold at all four zones except for /s/ (and to a much lesser extent for /r/) which turned out to be more variable at the pharynx than at the velar and palatal zones. On the other hand, the statistical results for the vowel data yielded a main effect of 'vowel' ($F(4, 195)=83.89$, $p< 0.001$) but no 'vowel' x 'zone' interaction meaning that, as shown in Figure 3, differences in area size for /u/ > /o/ > /a/ > /i, e/ apply equally to all four articulatory zones.

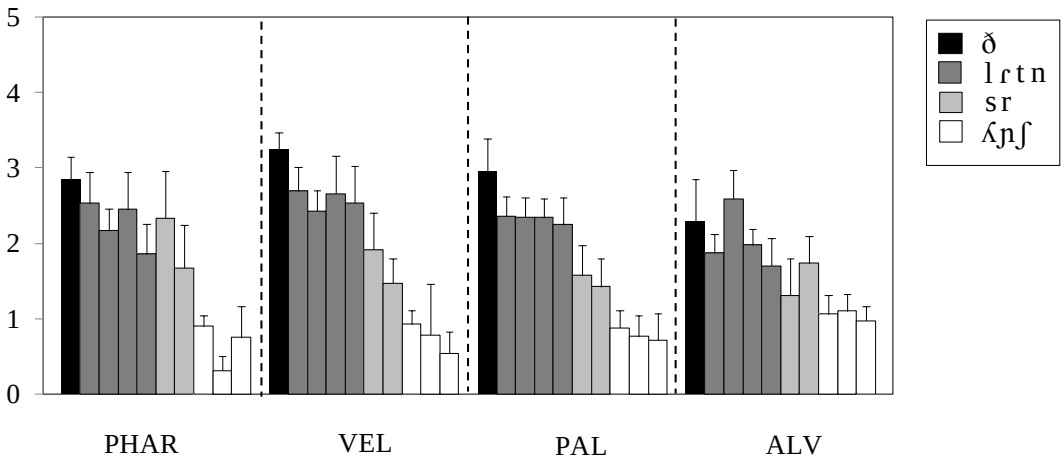


Figure 2. Cross-speaker normalized area values for consonants at the four articulatory zones ALV (alveolar), PAL (palatal), VEL (velar) and PHAR (pharyngeal). Error bars correspond to +/-1 standard deviation.

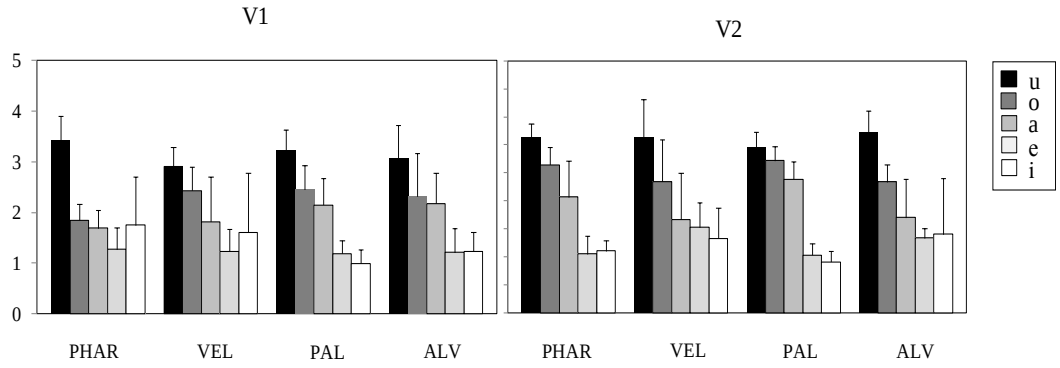


Figure 3. Cross-speaker normalized area values for vowels at the four articulatory zones ALV (alveolar), PAL (palatal), VEL (velar) and PHAR (pharyngeal). Error bars correspond to +/-1 standard deviation.

Discussion

Data reported in this study agree to a large extent with our initial hypothesis that coarticulatory resistance should vary in the progression /ʎ, ɲ, ʃ > /s, r/> /t, n, ɾ, l/ > /d/ ([ð]) for consonants and /i, e/ > /a/ > /o, u/ for vowels. Moreover, generally speaking, this hierarchy holds at the palatal, velar and pharyngeal zones where the tongue body is located and not just at the palatal zone, as reported by earlier EPG and EMA studies. Little contextual variability for palatal consonants and vowels (also for the trill /r/) at the three zones suggests that the entire tongue body is highly controlled during the production of these segmental units. Larger degrees of coarticulation were found to hold for the less constrained dentoalveolars /t, n, ɾ, l/ and for non-palatal vowels also at the palatal, velar and pharyngeal zones simultaneously. As for the highly constrained fricative /s/, there appears to be somewhat less coarticulatory variability at about constriction location than at the back of the vocal tract. These results accord with formant frequency data on coarticulatory resistance for the same consonants and vowels reported in the literature. They are also in support of the degree of articulatory constraint (DAC) model of coarticulation in that the extent to which a portion of the tongue body is more or less resistant to coarticulation depends both on its involvement in the formation of a closure or constriction and on the severity of the manner of articulation requirements.

Acknowledgments

This research has been funded by project FFI2013-40579-P from the Ministry of Innovation and Science of Spain, by ICREA (Catalan Institution for Research and Advanced Studies), and by the research group 2014 SGR 61 from the Generalitat de Catalunya.

References

Davidson, L. (2006) Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance, *JASA*, 120, 407-415.

Fitch, W. & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging, *JASA*, 106,1511–1522.

Mielke, J. (2015) An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons, JASA, 137, 2858-2869.

Recasens, D. (2014) *Fonètica i fonologia experimentals del català. Vocals i consonants* [*Experimental Phonetics and Phonology of the Catalan Language. Vowels and Consonants*] Institut d'Estudis Catalans, Barcelona.

Recasens, D. & Espinosa, A. (2009) An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan, *JASA*, 125, 2288-2298.

POSTERS

Wednesday December 9, 3:15-5:15

Tongue shape dynamics in swallowing

Mai Ohkubo[1], James M Scobbie[2]

1 Tokyo Dental College 2 CASL, Queen Margaret University

Introduction

During liquid swallowing, the tongue controls the liquid in a bolus in the oral cavity, changing shape, position and constriction to transport it into down into the pharynx. There are various methods for tongue movement measurement during swallowing: videofluoroscopy [Dodds et al (1990)], magnetic resonance imaging [Hartl et al (2003)] and ultrasound [Shawker et al (1983)]. Real-time ultrasound is simple, repeatable, and its dynamic soft tissue imaging may make it superior to others for swallowing research, and so we aim to test this hypothesis and measure certain spatial and dynamic aspects of the swallow in a consistent manner across participants.

Method

Eleven healthy adults (2 male and 9 female) between the ages of 19 and 35 participated in the study. Both thickened and thin liquids were used, and liquid bolus volumes of 10 and 25ml at room temperature were administrated to the subject using a cup. Three swallow tokens for each of the four bolus volume/viscosity were sampled, for a total of 12 swallows per subject.

The tongue surface was traced from the time at which the tongue moved up toward the palate at the start of swallowing, to the time when the entire tongue was in contact with the palate.

The distance (in mm) was calculated using AAA software, measuring along each radial fan line from the point where the tongue surface spline intersected the fan line to the point where the hard plate intersected the fan line in each individual plot. Each splines was calculated on sequential video frames while the middle of the tongue formed a concavity in preparatory positon. The depression distance was defined the longest distance from hard plate to tongue surface.

Results part 1

Qualitatively, there were differences between individual participants, and we defined quantitatively Measureable and Unmeasurable types. Figure 1 shows the most common type: Measureable, in which we could find a clear bolus depression in a cupped tongue surface. In 10ml thin liquids, we were able to find and measure the depression distance for all participants. In 10ml thickened liquids participants, we were not able to measure the depression distance for seven participants. Four participants were Unmeasurable in 25ml thickened liquids, and in 25ml thin liquids, two participants were Unmeasurable and one participant had unclear splines.

Results part 2

To make best use of the data, 10ml thin, 25ml thickened and 25ml thin (all Measurable types) were compared. Statistical comparison (ANOVA was possible therefore from 7 participants). The average maximum radial depression distance from palate to tongue surface was 20.9±4.3mm for 10ml thin liquid swallow compered 24.6±3.3mm for 25ml thin liquid swallow ($p < 0.001$). The average depression distance was 22.3±4.7mm for 25ml thickened liquid swallow compared with 25ml thin liquid swallow ($p < 0.01$).

Conclusion

We conclude that it is possible to use ultrasound tongue imaging to capture spatial aspects of swallowing. We will also discuss and exemplify dynamics of tongue constriction and the movement of the constriction from anterior to posterior.

References

Dodds W.J., Stewart E.T., Logemann J.A .Physiology and radiology of the normal oral and pharyngeal phases of swallowing, American Journal of Roentgenology 154(5):953-63,1990

Hartl D.M., Albiter M., Kolb .F et al.Morphologic parameters of normal swallowing events using single-shot fast spin echo dynamic MRI. Dysphagia,18(4): 255–62,2003

Shawker T.H., Sonies B, Stone M et al: Real-time ultrasound visualization of tongue movement during swallowing. J Clin Ultrasound, 11(9): 485–90,1983

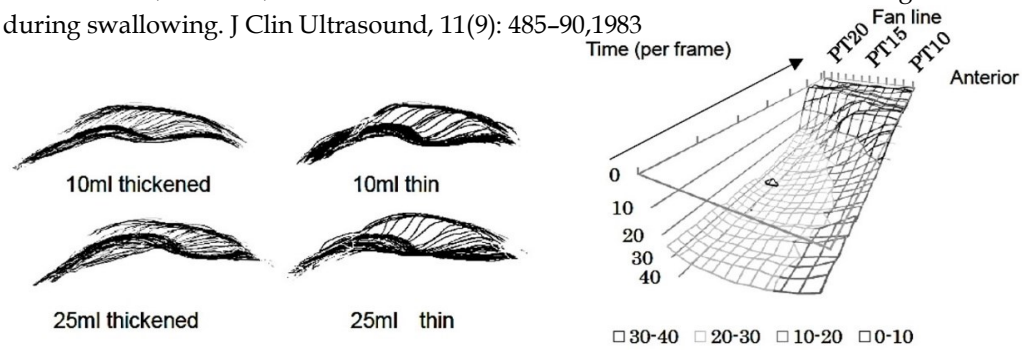


Figure 1. 22 year old female. Overlaid tongue curve splines (left) for four bolus types, and 3D time series (right) for the same 25ml thin bolus data, showing radial distance from tongue to palate along fan-shaped grid radii. The anterior constriction forms first at fanline PT10, then the contact spreads back across the palate to PT20. The anterior parts of the vocal tract are to the right in each image.

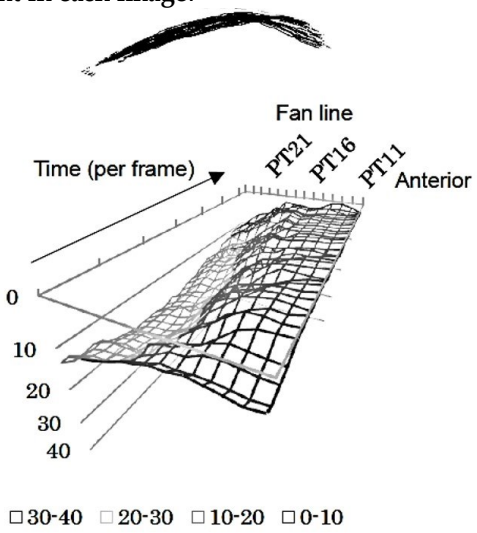


Figure 2

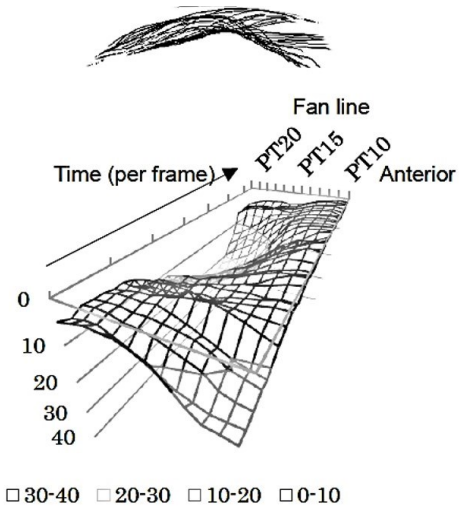


Figure 3

Figures 2 & 3 illustrate the Unmeasurable types. Figure 2 is a 19 year old female in which the tongue's surface didn't make a travelling concavity and the detected movement was only very slight. Figure 3 shows data from a 24 year old female with an anterior concavity at the start and a dorsal concavity later (just before, at the end of the transport, the near-complete closure), but, in between these times, the front / middle of the tongue didn't form the clear concavity travelling in a posterior direction as might be expected. This may be because, unusually, she held the dorsal part of her tongue near to or touching the palate at the start of the process.

Recordings of Australian English and Central Arrernte using the EchoBlaster and AAA

Marija Tabain (*La Trobe University, AUSTRALIA*)
Richard Beare *(Monash University, and Murdoch Children's Research Institute, AUSTRALIA)*

We recently recorded seven speakers of Australian English, and seven speakers of Central Arrernte, a language of Central Australia, using the Telemed Echo Blaster 128 CEXT-1Z, the Articulate Instruments stabilization helmet, the Articulate Instruments pulse-stretch unit, and the AAA software version 2.16.07. In addition we used an MBox2 Mini soundcard, a Sony lapel microphone (electret condenser ECM-44B), and an Articulate Instruments Medical Isolation Transformer. Typical frame rate was 87 f.p.s., using a 5-8 MHz convex probe set to 7 MHz, a depth of 70 mm and a field of view of 107.7 degrees (70%).

The recordings of Australian English served primarily as practice before taking the equipment to Central Australia for field recordings. Many problems were initially encountered, particularly regarding synchronization, and this required bug fixes to the software. Data from one speaker was entirely discarded, and other speakers had sporadic synchronization problems.

For both the English and the Arrernte recordings, one speaker of each language did not display a visible contour outline for the tongue – in the case of Arrernte, this speaker was simply not recorded, since we had ended up discarding the data from the English speaker who displayed this particular characteristic. For each language, about 2-3 speakers displayed good tongue contour outlines; the remaining speakers have slightly less clear outlines. The English speakers' data have been tracked using the AAA software, with manual corrections where needed.

Both WAV and Spline data for English have been exported from AAA and read into the EMU speech analysis system, interfaced with the R statistical package. Simple plotting routines have been successfully conducted on the English data, which focused on hVd, hVl and lVp sequences of English (i.e. effects of preceding vs. following laterals on the various vowels of Australian English). Tongue contours have been plotted across time for a given token, and also at the temporal midpoint for a given set of tokens. We plan to present these preliminary English results in Hong Kong.

The effects of blindness on the development of articulatory movements in children

Pamela Trudeau-Fisette, Christine Turgeon, Marie Bellavance-Courtemanche, And Lucie Ménard

Laboratoire de phonétique, Université du Québec à Montréal, Montréal, Canada

INTRODUCTION

It has recently been shown that adult speakers with congenital visual deprivation produce smaller displacements of the lips (visible articulator) than their sighted peers (Ménard et al., 2013). As a compensatory maneuver, blind speakers move their tongue more than sighted speakers. Furthermore, when vowels are produced under narrow focus, a prosodic context known to enhance distinctiveness, blind speakers mainly alter tongue movements to increase perceptual saliency, while sighted speakers alter tongue and lip movements (Ménard et al., 2014). However, from a developmental perspective, not much is known about the role of blindness in speech production. The objective of this paper was therefore to investigate the impact of visual experience on the development of the articulatory gestures used to produce intelligible speech.

METHOD

Eight congenitally blind children (mean age: 7 years old; range: 5 to 11 years) and eight sighted children (mean age: 7 years old; range: 5 to 11 years) were recorded while producing repetitions of the French vowels /i/, /a/, and /u/ in a /bVb/ sequence in two prosodic conditions: neutral and under contrastive focus. The prosodic contexts were used here to manipulate distinctiveness and elicit hyperarticulation. Lip and tongue movements, as well as the acoustic signal, were recorded using a SONOSITE 180 ultrasound system and a video camera. The current paper focuses on acoustic measures and lingual measurements. Formant frequencies, fundamental frequency values, and tongue shapes (Li et al., 2005) were extracted at vowel midpoint. Measures of curvature degree and asymmetry (tongue shape) were extracted following Ménard et al.'s (2012) method.

RESULTS

Preliminary analyses of the data show that blind children move their tongue to a greater extent than their age-matched sighted peers. Trade-offs between lip and tongue displacements, inferred from acoustic measurements, are discussed. Overall, our results show that blindness affects the developmental trajectory of speech.

REFERENCES

Li, M., Kambhamettu, C., and Stone, M. (2005).“Automatic contour tracking in ultrasound images,” Clin. Ling. and Phon., 19, 545–554.

Ménard, L., Aubin, J., Thibeault, M., and Richard, G. (2012). “Comparing tongue shapes and positions with ultrasound imaging: A validation experiment using an articulatory model*,” Folia Phoniatr. Logop.* 64, 64-72.

Ménard, L., Toupin, C., Baum, S., Drouin, S., Aubin, J., and Tiede, M. (2013). “Acoustic and articulatory analysis of French vowels produced by congenitally blind adults and sighted adults,” *J. Acoust. Soc. Am.* 134, 2975-2987.

Ménard, L., Leclerc, A., and Tiede, M. (2014): "Articulatory and acoustic correlates of contrastive focus in congenitally blind adults and sighted adults", *Journal of Speech, Language, and Hearing Research,* 57, 793-804.

POSTER

An EPG + UTI study of syllable onset and coda coordination and coarticulation in Italian

Cheng Chen, Chiara Celata, Irene Ricci, Chiara Bertini and Reza Falahati

Scuola Normale Superiore, Pisa, Italy

1. Introduction

This study is concerned with the methodological challenges of studying articulatory coordination of onset and coda consonants by means of an integrated system for the acquisition, real-time synchronization and analysis of acoustic, electropalatographic and ultrasound data. Electropalatographic panels (EPG) are responsible for the contact (closure/aperture) between the tongue and palate, while ultrasonographic (UTI) images provide the complementary information of the sagittal profiles of tongue synchronised with EPG data during the articulation of consonants and vowels in the speech chain. This original system makes it possible to process simultaneously the information of both linguo-palate contact and the movement of tongue for reaching its target (Spreafico et al. 2015).

The system is used to capture simultaneous data on linguo-palatal contact and tongue sagittal profiles of /s/, /l/ and /k/ adjacent to /a/ and /i/ as produced by native speakers of a Tuscan variety of Italian.

Using EPG and UTI data to investigate temporal and spatial coordination of consonant-vowel sequences is challenging to the extent that the identification of 'anchor points' for temporal measurements is not straightforward starting from information about whole tongue or tongue-palate configurations (or at least, not as straightforward as when starting from trajectories of points, as in EMA-based studies). At the same time, the two-channel experimental environment provides fine-grained spatial, in addition to temporal, information, namely, by allowing the analysis of coarticulatory activity for the selected anchor points and for the temporal lags between them.

The poster will illustrate the innovative audio-EPG-UTI synchronization system and offer some preliminary considerations about the methodological challenges related to the investigation of temporal and spatial coordination patterns in onset and coda consonants.

2. Motivation of the study

According to the articulatory model of syllable structure, the temporal and spatial coordination of articulatory gestures is conditioned by position in the syllable. The onset consonants are supposed to be more stable and to have a greater degree of constriction with respect to coda consonants (syllabic asymmetry; Krakow 1999). Moreover, the (temporal) coordination between an onset singleton consonant and the following nuclear vowel is found to be more stable than that between a nuclear vowel and the coda consonant (Browman and Goldstein 1988, 2000).

Although the stability of onset consonants is confirmed by many a study in the last ten years, recent research has revealed that the onset-nucleus coordination is also predicted by the articulatory properties of the consonant (e.g. Pastaetter & Poulier 2015); specifically, it is modulated by the degree of coarticulation resistance (Recasens and Espinosa 2009) of the consonant involved.

Such phenomena suggest that the intrinsic articulatory property of consonant might influence the temporal (and spatial) coordination between articulatory gestures. Cross-linguistic comparisons are expected to provide more evidence about the supposed interaction between coarticulatory patterns and gestural timing.

3. Description of the experiment

For this study on Italian, the corpus is composed of 12 stimuli, all disyllabic pseudo-words or very infrequent words. Each stimulus is inserted in a carrying sentence providing the same segmental context in which a bilabial consonant for all stimuli. The target consonants are /s/, /l/ and /k/; according to the DAC

model (e.g. Recasens & Espinosa 2009), they have a high, intermediate and low degree of coarticulatory resistance, respectively. These consonants are analyzed both as onsets and as codas, i.e. in CV and VC contexts. The V is /a/ in one series, /i/ in another series. The stimuli with /a/ are produced twice: first in a prosodically neuter condition, then in a prosodically prominent position in which the target word bears a contrastive pitch accent.

Table 1 provides an example of carrying sentences and the list of stimuli. In the carrying sentence, the first repetition of the target stimulus corresponds to the prosodically neuter condition, while the second corresponds to the prosodically prominent condition (contrastive pitch accent). Following the hypothesis that laryngeal and supralaryngeal gestures tend to be coordinated (e.g. Ladd 2006, Muecke et al. 2012), we expect that also prosodic prominence can influence the way in which the onset-coda contrast is realized, either by enhancing or by reducing it.

Carrying-sentences	Pronuncia saba molte volte. ("He pronounces saba a lot of times.") Pronuncia seba? No, pronuncia SABA molte volte! ("Does he pronouce seba? No, he pronounces SABA a lot of times!")					
	/s/		/l/		/k/	
	CV	VC	CV	VC	CV	VC
/a/	Saba	bass	laba	bal	capa	pac
/i/	Siba	bis	liba	bill	kipa	pic

Table 1. Example of carrying sentences and list of the 12 target words in the corpus

The recordings were made in the phonetics laboratory of Scuola Normale Superiore, Pisa. Ultrasound data were captured using a MindRay device with a acquisition rate of 60 Hz, an electronic micro-convex probe (Mindray 65EC10EA 6.5 MHz) and a stabilization headset; electropalatographic data were captured via the WinEPG system by Articulate Instrument (SPI 1.0) recording palate images at 100 Hz; EPG, UTI and audio data were acquired and real-time synchronized using the Articulate Assistant Advanced (AAA) software environment and a video/audio synchronization unit. Two digital tones were produced and used to synchronize both EPG and UTI signals with the audio signal.

4. Methodological challenges

The two-channel synchronized articulatory approach allows the analysis of the temporal coordination of gestures *and* the coarticulatory patterns underpinnings gestural coordination in *one* output. For such goal to be fulfilled, it is however necessary to define a series of temporal landmarks allowing the estimation of gestures' relative distance (temporally and spatially).

Consonants and vowels are manually segmented according to the inspection of waveform and spectrogram (after exportation into Praat). In each vocalic or consonantal interval it is then possible to locate time-points for, respectively, the vocalic anchor and the reaching of maximum consonantal constriction. The vocalic anchor is the point in which the vowel reaches its target configuration (i.e., maximal predorsum lowering and tongue flattening for /a/, maximal predorsum raising for /i/). The maximum consonantal constriction is the time-point in which the articulatory target is reached (i.e. maximum constriction in the relevant lingual and palatal areas and minimal influence of V-to-C coarticulation). These two points are taken as references for the calculation of intergestural timing (or temporal distance, measured in ms) and of the coarticulatory modification of C (or spatial distance, measured in terms of changes in EPG indices, formant values and lingual profiles) as a function of V quality changes, position in the syllable and prosodic prominence.

To locate the V anchor and the maximum C constriction point, the EPG and UTI outputs for the selected acoustic intervals are first independently evaluated. E.g. for a /li/ stimulus, according to tongue profile qualitative inspection, the stable maximal constriction for /l/ is defined as the sequence of UTI frames showing apical raising and contextual dorsum flattening, before the anterodorsum fronting caused by the anticipation of the gesture for the /i/. The relevant UTI interval is labeled Δt1. Similarly, according to

which there is maximum anterior constriction (with partial lateral contact) and before lateral obstruction and dorsum raising (also for anticipatory coarticulation). The relevant EPG interval is labeled Δt2. As a subsequent step the extension of Δt1 and Δt2 is simultaneously evaluated. The first temporal instant that falls within both Δt1 and Δt2 intervals corresponds to the maximum C constriction time-point. The V anchor time-point for /i/ is identified according to the same procedure, within the acoustic interval of the vocalic nucleus.

The temporal coordination of the consonantal and vocalic gestures in the different syllabic contexts (CV vs VC) can then be evaluated in conjunction with the spatial coarticulatory coordination for the two gestures. The study also allows the analysis of the effects of coarticulatory resistance (as evaluated from the comparison of the three consonants in the /a/ vs. /i/ context) and of prosodic prominence (as evaluated from the comparison between the prosodically neuter and the pitch accent condition) on C and V gestural organization.

Browman, C.P. & Goldstein, L. (1988). Some notes on syllable structure in Articulatory Phonology. *Phonetica* 45, 140-155.

Browman, C.P. & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée*, 5, 25–34.

Krakow, R. A. (1999). Physiological organization of syllables: A review. *Journal of Phonetics* 27. 23–54.

Ladd, D. R. (2006). Segmental anchoring of pitch movements: Autosegmental association or gestural coordination? *Italian Journal of Linguistics* 18(1), pp. 19–38.

Mücke, D., Nam, H., Hermes, A. and L. Goldstein (2012). Coupling of tone and constriction gestures in pitch accents. In *Consonant Clusters and Structural Complexity*, Mouton de Gruyter, pp. 205-230.

Pastaetter M. & M. Pouplier (2105) Onset-vowel timing as a function of coarticulation resistance: Evidence from articulatory data. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, 10-14 August 2015.

Recasens D. & A. Espinosa (2009) An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *Journal of the Acoustical Society of America* 125, 2288–2298.

Spreafico L., C. Celata, A. Vietti, C. Bertini & I. Ricci (2015) An EPG+UTI study of Italian /r/. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, 10-14 August 2015.

A Kinect 2.0 system to track and correct head-to-probe misalignment

Sam Johnston[1], **Diana Archangeli**[1,2], and **Rolando Coto**[1]

[1]*Department of Linguistics, University of Arizona*
[2]*Department of Linguistics, University of Hong Kong*

In ultrasound experimentation, a constant alignment between a subject's head and the ultrasound probe is essential to a valid analysis. This fixed head-to-probe alignment is critical to obtain accurate ultrasound images of the tongue that can be reliably compared with one another. Consequently, there has been much work to develop an effective method of securing a subject's head in relation to the ultrasound probe. Previous methods have included the HATS system (Stone and Davis, 1995), the use of a fitted helmet (McLeod and Wrench, 2008), and more recently a elastic strap (Derrick et al., 2015), all of which use a physical apparatus to manually fix the head-to-probe alignment. Two additional systems, the Palatoglossatron (Baker, 2005), and HOCUS/OptoTrak (Whalen et al., 2005) are systems which track the position of the head, instead of immobilizing the head. These each require the subject to wear additional equipment.

One limitation of the Palatoglossatron (Baker, 2005), is that it is primarily intended to correct for pitch-dimension misalignment, and does not address the dimensions of yaw and roll. HOCUS (Whalen et al., 2005) requires infrared diodes to be placed on a tiara or directly onto the head to track its possible movement and misalignment. Yet these diodes themselves are subject to possible movement during the experiment (cf. Roon et al., 2013), throwing off head tracking.

The current study utilizes the Kinect 2.0 head-tracking API (Han et al., 2013) to identify and track the the location of a head in 3D space in real time. This system allows for free head movement and also does not require any special devices to be worn, and therefore is completely non-invasive, making it particularly suitable for young children and elderly subjects. The Kinect has been integrated into a custom-designed system that will alert subject and researcher when the subject's head becomes misaligned from a stationary ultrasound probe.

The purpose of the present study was to establish the accuracy of the Kinect's head-tracking measurements. Video cameras were placed to the side, in front of, and above the subject during the experiment, capturing the angle of the head in each dimension of pitch, yaw, and roll as it moves from center. Images from the videos were taken, and the measurements of the Kinect system were verified by hand-measuring the video images.

Results indicate that the Kinect's tracking of head movement is quite similar for each of pitch, yaw, and roll. For each of these dimensions, Whalen et al. (2005) describes acceptable ranges of head movement which does not significantly alter the quality of an ultrasound image. They find that for any dimension, 5 degrees of movement is tolerable. In the present study, when the (hand-measured) head-tilt was within 5 degrees in either direction, the Kinect's measurement values diverged no more than 2 degrees from the hand-measured angle. This demonstrates that the Kinect head-tracking software can be used to set limits that will conservatively keep the subject's head within an acceptable range of movement.

References

Baker, A. (2005). Palatoglossatron 1.0. *University of Arizona Working Papers in Linguistics*.

Derrick, D., Best, C., and Fiasson, R. (2015). Non-metallic ultrasound probe holder for co-collection and co-registration with ema. pages 1–5.

Han, J., Shao, L., Xu, D., and Shotton, J. (2013). Enhanced computer vision with microsoft kinect sensor: A review. *IEEE Trans. Cybernetics*, 43(5).

McLeod, S. and Wrench, A. (2008). Protocol for restricting head movement when recording ultrasound images of speech. *Asia Pacific Journal of Speech, Language, and Hearing*, 11:23–29.

Roon, K., Jackson, E., Nam, H., Tiede, M., and Whalen, D. H. (2013). Assessment of head reference placement methods for optical head-movement correction of ultrasound imaging in speech production. *Journal of the Acoustical Society of America*, 134:4206.

Stone, M. and Davis, E. P. (1995). A head and transducer support system for making ultrasound images of tongue/jaw movement. *Journal of the Acoustical Society of America*, 98:3107–3112.

Whalen, D. H., Iskarous, K., Tiede, M., Ostry, D., 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.

Poster

Title: Articulatory Settings of Japanese-English Bilinguals

Authors: Ian Wilson, Yuki Iguro, Julián Villegas

Affiliation: University of Aizu, Japan

Abstract: In a similar experiment to Wilson & Gick (2014; JSLHR), who investigated the articulatory settings of French-English bilinguals, the present study is focused on Japanese-English bilinguals of various proficiencies. We analyze interspeech posture (ISP), and look at the differences between individuals and whether this is correlated with the perceived nativeness of the speakers in each of their languages.

The UltraPhonix Project: Ultrasound Visual Biofeedback for Heterogeneous Persistent Speech Sound Disorders

Joanne Cleland[1], James M. Scobbie[2], Zoe Roxburgh[2] and Cornelia Heyde[2]

[1]University of Strathclyde, Glasgow [2]Queen Margaret University, Edinburgh.

Ultrasound Tongue Imaging (UTI) is gaining popularity as a visual biofeedback tool that is cost-effective and non-invasive. The evidence for Ultrasound visual biofeedback (U-VBF) therapy is small but promising, with around 20 case or small group studies. However, most studies originate from the USA and Canada, and focus on the remediation of delayed/disordered /r/ production (for example McAllister et al., 2014). While ultrasound is ideal for visualising /r/ productions, it also offers the ability to visualise a much larger range of consonants and all vowels, for example Cleland et al. (2015) report success in treating persistent velar fronting and post-alveolar fronting of /ʃ/. This paper will report on a new project, "UltraPhonix" designed to test the effectiveness of U-VBF for a wider range of speech sounds in more children than previously reported.

The UltraPhonix project will recruit 20 children aged 6 to 15 with persistent speech sound disorders affecting vowels and/or lingual consonants in the absence of structural abnormalities. Since the children will have a range of different speech targets, the project design is a single-subject, multiple baseline design, with different wordlists (probes) designed according to the presenting speech error. Children will receive 10 sessions of U-VBF therapy, preceded by three baseline probes, and followed by two maintenance measures. This project uses a high-speed Ultrasonix SonixRP machine running Articulate Assistant Advanced software (Articulate Instruments, 2012) at 121 frames per second allowing us to capture dynamic information about the children's speech errors for diagnostic purposes. Moreover, the ultrasound probe is stabilised with a headset, allowing us the unique capability to compare ultrasound data across assessment and therapy sessions (see Cleland et al., 2015). Bespoke U-VBF therapy software has already been designed allowing us to super-impose hard palate traces on the ultrasound image and view target videos of typical speakers articulating the target speech sounds.

Our poster presents the methodology of our new project and give sample data from the first group of participants recruited to the project.

References

Articulate Instruments Ltd 2012. *Articulate Assistant Advanced User Guide: Version 2.14*. Edinburgh, UK: Articulate Instruments Ltd.

Cleland, J., Scobbie, J.M. & Wrench, A., (2015). Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics and Phonetics*. Pp. 1-23.

Poster

McAllister Byun, T. M., Hitchcock, E. R., & Swartz, M. T. (2014). Retroflex versus bunched in treatment for rhotic misarticulation: Evidence from ultrasound biofeedback intervention. *Journal of Speech, Language, and Hearing Research, 57(6*), 2116-2130.

Gradient Acquisition of Velars via Ultrasound Visual Biofeedback Therapy for Persistent Velar Fronting.

Joanne Cleland[1], James M. Scobbie[2], Jenny Isles[1], Kathleen Alexander[2]

[1]University of Strathclyde, Glasgow [2]Queen Margaret University, Edinburgh.

BACKGROUND: Velar fronting (substituting /k,g,ŋ/ with [t,d,n] is a well attested phonological process in both the speech of young typically developing children and older children with speech sound disorders, with typically developing children acquiring velars by the time they are three and half years old. This particular speech error is of interest because absence of velars in the phonetic inventory at three years of age is predictive of phonological disorder and children who fail to differentiate coronal (tongue tip) and dorsal (tongue body/back) articulations may present with motoric deficits. When children fail to acquire velars in the process of normal development, speech therapy techniques which draw children's attention to the homophony in their speech sound systems is can be effective. However, a subset of children become persistent velar fronters, still unable to articulate velar consonants well into the school years. Cleland et al. (2015) showed that it is possible to remediate persistent velar fronting using Ultrasound Visual Biofeedback (U-VBF), but like most studies of instrumental articulatory therapies, very little about how the children acquire the new articulation is known, with most studies presenting pre and post therapy assessment data only.

This paper presents data from multiple assessment time-points from the Cleland et al. (2015) study. Given that these children may have a motoric deficit it is important to look at the fine phonetic detail of their articulations in order to identify how they begin to make new articulatory gestures and how these gestures change over time.

METHOD: Data from four children with persistent velar fronting was analysed. Each child received 12 sessions of therapy with U-VBF and five assessment sessions. All ultrasound data was recorded with a high-speed Ultrasonix SonixRP machine running Articulate Assistant Advanced software (Articulate Instruments, 2012) at 121 frames per second. The probe was stabilised with a headset and data was normalised across sessions using hard-palate traces. Attempts at velar and alveolar minimal pairs from pre-therapy, mid-therapy, post-therapy and six weeks post therapy were annotated at the burst. The nearest ultrasound frame to the annotation point was selected and a spline indicating the tongue surface fitted to the image using the automatic function in AAA software. We calculated radially "kmax-t" where "kmax" was the tongue spline point at the burst of /k/ further from the probe and "t" was the tongue spline point along the same fan line. Results were compared to those for 30 typical children. In addition, we used the methodology from Roxburgh et al. (under revision) to perceptually evaluate the children's attempts at words containing velars at four of the time points.

RESULTS: Three of the children achieved a dorsal articulation after only three sessions of U-VBF. One child (05M) achieved no velars after 12 sessions of therapy, but went on to achieve velar stops after a second block of U-VBF. In each child, pre-therapy kmax-t was near zero, indicating no difference in tongue shapes for /t/and /k/ and suggesting no covert contrast. Mid-therapy, two children overshot the optimum kmax-t (heard as uvular) and subsequently moved in a gradient fashion towards kmax-t in the normal range. The other two children had kmax-t small than normal at mid-therapy, but increased this measurement to normal levels six weeks post-therapy. Results of the perceptual experiment show similarly gradient improvement, with listeners rating later attempts at words

containing velars as more like those of adults, even when phonetic transcription rated adjacent session recordings as both 100% on target. This gradual improvement in the articulation if velars suggests a motor-based deficit in these children with persistent velar fronting.

References.

Cleland, J., Scobbie, J. M., & Wrench, A. A. (2015). Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical linguistics & phonetics*, (0), 1-23.

A non-parametric approach to functional ultrasound data: A preliminary evaluation

Alessandro Vietti*, Alessia Pini°, Simone Vantini°, Lorenzo Spreafico*, Vincenzo Galatà*

**** Free University of Bozen, ° MOX - Department of Mathematics, Politecnico di Milano***

In the last decades, functional data analysis (FDA) techniques have been successfully applied to the analysis of biologic data. Some recent examples pertain to the analysis of blood vessel shapes (Sangalli et al., 2014), proteomic data (Koch at al, 2014), human movements data (Ramsay et al., 2014), and neural spike-trains (Wu et al., 2014).
The aim of the present study is to apply FDA techniques to a data set of tongue profiles. In detail, we carry out a comparison of two alternative methods that could be suited for the analysis of tongue shapes, namely smoothing spline ANOVA (SSANOVA) (Gu 2002; Davidson 2006) and the interval-wise testing (IWT) (Pini&Vantini, 2015). The two techniques basically differ in the inferential process leading to the construction of confidence intervals. SSANOVA is indeed a parametric approach based on Bayesian inference. On the contrary, IWT is a non-parametric approach based on permutation tests. In particular, IWT neither assumes data to follow a Gaussian distribution, nor needs to specify any a-priori information about the parameters defining the Gaussian distribution.

The two techniques are applied to a dataset of tongue shapes recorded for a study on Tyrolean, a German dialect spoken in South Tyrol (Vietti&Spreafico 2015). In detail, data are composed of 160 tongue profiles of five variants of uvular /r/ recorded from one native speaker of Tyrolean (F, 33 y.o.). The five groups of curves corresponds to five different manners of articulation: vocalized /r/, approximant, fricative, tap, and trill.
Firstly, SSANOVA is performed following the standard procedure presented in Davidson (2006), using the *gss* R package and the *ssanova* function (Fig 1. on the left). Smoothing spline estimate and Bayesian confidence interval for comparison of the mean curves are obtained as well as the interaction curves with their relative confidence intervals.
Secondly, the IWT is performed. The IWT provides two kinds of outputs:

1) Non-parametric 95% confidence bands for the position of the tongue within the five groups (Fig. 1)
   Non-parametric point-wise (angle-wise) confidence bands are estimated for the mean position of the tongue within each of the five groups. The confidence bands are estimated, for each point of the domain, by means of non-parametric permutation techniques (Pesarin, 2010), with a confidence level of 95% (Fig. 1 on the right).

2) Non-parametric interval-wise tests for group comparisons (Fig. 2)
   We test the equality of the functional distributions of each pair of groups. All tests are based on the IWT proposed in Pini&Vantini (2015) which - differently from the SSANOVA - is able to identify the regions of the domain presenting significant differences between groups, by controlling the probability of wrongly selecting regions with no-difference. The procedure results in the evaluation of an adjusted p-value function that can be thresholded to select the regions of the domain presenting significant differences. Such selection is provided with a control of the interval-wise error rate.

From a preliminary evaluation, the two techniques represent the differences among the five groups of functions in a very similar way when the sample size is sufficiently large, but differently if the sample size is low and the curve distribution is far from being Gaussian. A number of other critical issues emerges from the comparison that deserves further investigation.
In particular the following ones will be discussed.

a) SSANOVA results turn out to be extremely sensitive with respect to the choice of the B-spline basis chosen to model the curves. This is due to the fact that in the SSANOVA the generative probabilistic model is directly built on the coefficients of the basis expansion and not on the curves themselves.
b) SSANOVA results – coherently with the Bayesian perspective – could be strongly dependent on the prior distribution. This fact, for groups with a reduced sample size, leads to confidence bands not centered on the corresponding groups of curves.
c) Within each group the permutation confidence bands seem to better recover the different point-wise variability observed along the tongue profiles.
d) IWT allows group comparisons in terms of adjusted p-value functions, which may result in a more informative and detailed representation of the regions of the tongue where a significant difference is located (especially in the pairwise scatter-matrix representation Fig. 1).

A further speculation may arise from points (a,b): the ITW approach seems to be more stable and more tolerant to unbalanced design or at least to groups (r-variants) characterized by a small number of observations. The computational stability in case of unbalanced design should be more carefully investigated in order to evaluate which technique could be applied to more "naturalistic" data coming for instance from non-experimental settings.

References:

Davidson, L. (2006), Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance, JASA 120 (1), 407-415.

Gu, C. (2002), *Smoothing Spline ANOVA models*. Springer, New York.

Koch, I., Hoffmann, P., and Marron, J. S. (2014). Proteomics profiles from mass spectrometry. Electronic Journal of Statistics, 8(2), 1703-1713.

Pesarin, F. and Salmaso, L. (2010). *Permutation tests for complex data: theory, applications and software*. John Wiley & Sons Inc, Chichester.

Pini, A. and Vantini, S. (2015), Interval-wise testing for functional data. Technical Report **30/2015**, MOX – Department of Mathematics, Politecnico di Milano.

Ramsay, J. O., Gribble, P., and Kurtek, S. (2014). Description and processing of functional data arising from juggling trajectories. *Electronic Journal of Statistics*, *8*(2), 1811-1816.

Sangalli, L. M., Secchi, P., and Vantini, S. (2014). AneuRisk65: A dataset of three-dimensional cerebral vascular geometries. Electronic Journal of Statistics, 8(2), 1879-1890.

Vietti, A. and Spreafico, L. (2015), An ultrasound study of the phonetic allophony of Tyrolean /r/, ICPhS 2015 Proceedings.

Wu, W., Hatsopoulos, N. G., and Srivastava, A. (2014). Introduction to neural spike train data for phase-amplitude analysis. *Electronic Journal of Statistics*, *8*(2), 1759-1768.

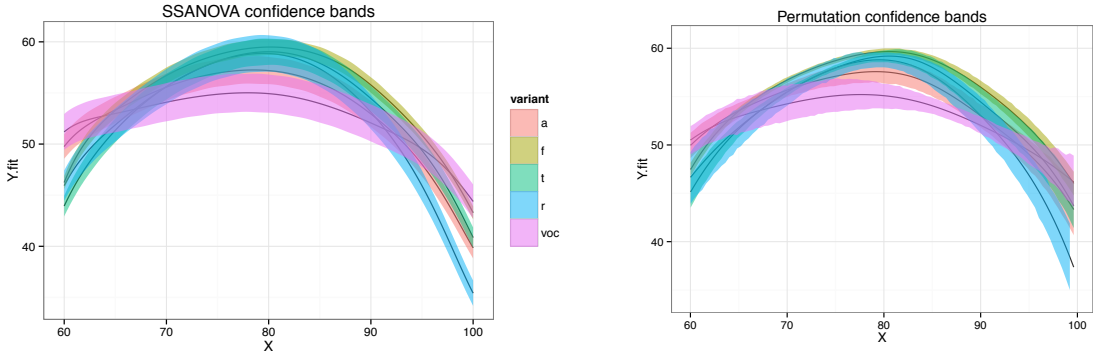


Figure 1.
Confidence bands for the five groups of tongue profiles obtained via SSANOVA (left) and permutation bands (right).

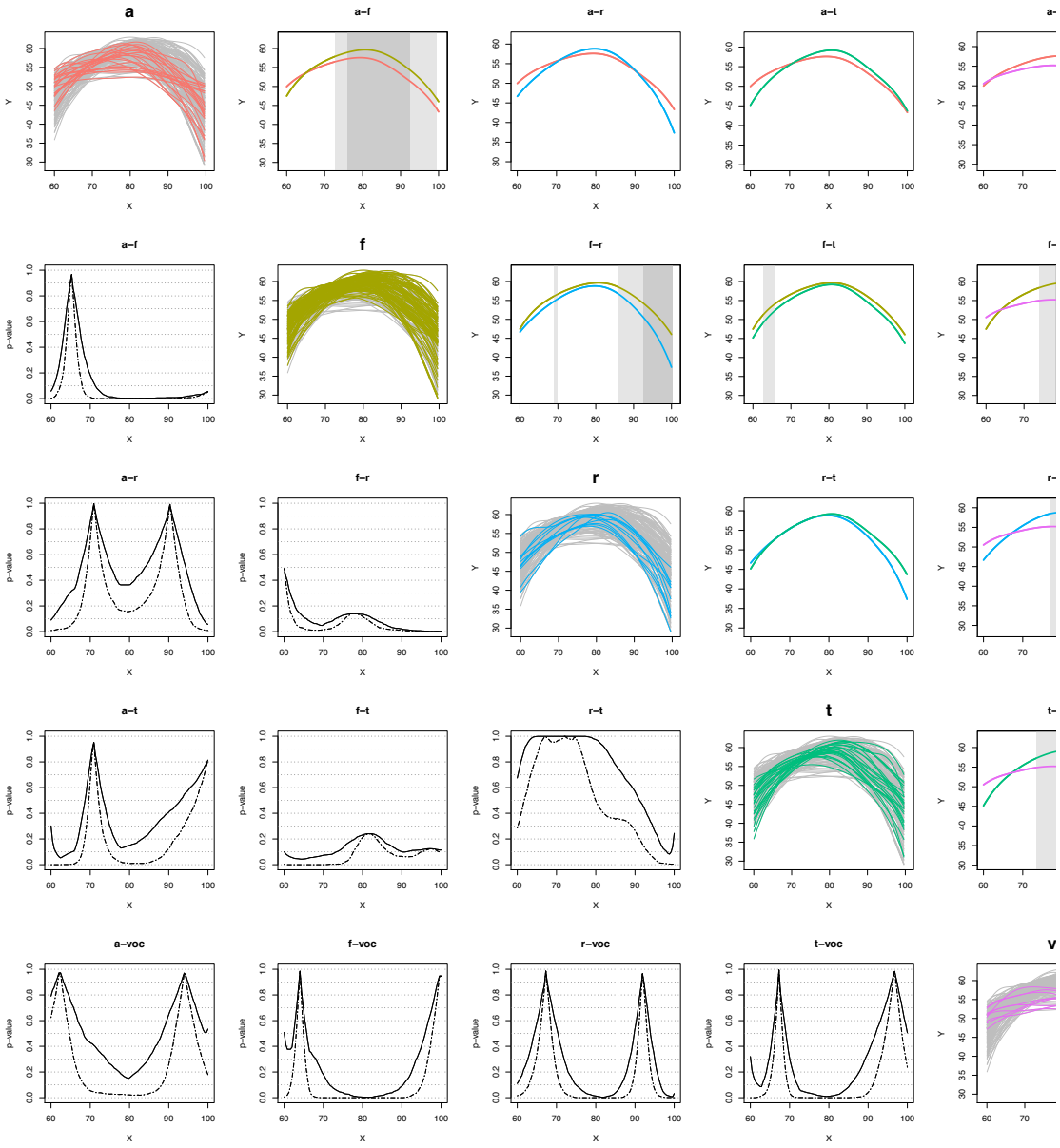


Figure 2.
Pairwise scatter-matrix of two-group comparisons obtained via the IWT procedure. Diagonal panels: tongue profiles of the five groups. Lowed diagonal panels: adjusted (full line) and unadjusted (dashed lines) *p*-value functions. Upper diagonal panels: means of the compared groups and gray areas representing significantly different intervals at 1% (dark gray) and 5% (light gray) significance levels.

Effects of phrasal accent on tongue movement in Slovak

Lia Saki Bučar Shigemori
Marianne Pouplier
Štefan Beňuš

This study examines the effect of phrasal accent on tongue movement for vocalic and consonantal nuclei in Slovak using ultrasound.

The main difference between vowels and consonants is grounded in their syllabic affiliation in that vowels always occupy the nuclear position while consonants occupy the onset or coda position. Prosody is another domain that divides vowels from consonants, in that, broadly speaking, vowels carry the prosodic and consonants the lexical information. Slovak has two syllabic consonants, /l/ and /r/, which can also occupy the nucleus of a stressed syllable. This enables us to examine the implementation of phrasal accent on vowels and consonants in a lexically stressed nucleus, the position where prosodic effects are expected to be most prominent.

Previous research has revealed two strategies on how prosodic prominence is produced. The first is `sonority expansion`, which is achieved by expanding the oral cavity, usually by lowering jaw and tongue (Beckman et al., 1992). The second one is `hyperarticulation` (De Jong, 1995). For many vowels, these two strategies go by and large hand in hand because hyperarticulation would lead to an even wider opening of the oral cavity, which would also enhance sonority. For consonants, on the other hand, hyperarticulation would predict a tighter constriction, which requires a movement opposite to what would be required for sonority expansion.

In the current paper we want to examine whether phrasal accent is implemented on consonantal nuclei as it is on vowels. We analyze the nucleus of the first syllable of the two phonologically valid nonsense words *pepap* (vocalic nucleus /e/) and *plpap* (consonantal nucleus /l/). Word stress in Slovak is fixed on the first syllable. Fundamental frequency is a robust indicator for phrasal accent in Slovak (Král', 2005) and was used to control whether speakers correctly produced the phrasal accent. The two target words were inserted in two carrier phrases to elicit the two accent patterns:

Accented targetword
Pozri, ved' on mi ***pepap*** dal.
(Look, he even gave me pepap.)

Unaccented targetword
Pozri, aj **Ron** mi *pepap* dal.
(Look, also Ron gave me pepap.)

To see whether the implementation of phrasal accent can be observed on vocalic, as well as consonantal nuclei, we first want to examine whether there are:

1. Differences between the F1 and F2 movement throughout the nucleus,
2. Differences in tongue contours at the beginning, midpoint and endpoint of the nucleus

for the two accent patterns, separately for vowels and consonants.

Slovak has a dark /l/, which consists of two gestures: the *consonantal* tongue tip movement and the *vocalic* tongue back movement (Sproat and Fujimura, 1993). If prosody is to be carried by vowels, we expect weaker tongue tip constriction in the accented position and a more prominent retraction of the tongue body. To test whether prosody is carried only by vowels, we want to look at

1. Whether the tongue tip constriction is present,
2. Whether the tongue tip constriction is present only at the beginning or end of the nucleus,
3. Whether both gestures are influenced by accentuation if they are present.

We present acoustic and articulatory data for one speaker. In Figure 1 the movement of F1 and F2 throughout the target nucleus are visualized. Figures 2 and 3 show the tongue contours at the beginning, midpoint and end of the two target nuclei. The nucleus has been defined acoustically, starting with the beginning of voicing after the burst of the preceding /p/ and ending with the closure for the following /p/.

We see accent induced contrasts for vowels and consonants in the formant movement as well as the tongue contours.

For the vocalic nucleus, F1 is flat with a slight fall towards the end for the unaccented condition. F2 has a shorter flat part followed by a steeper fall and is overall lower for the unaccented condition. The tongue contours are slightly further back for the unaccented condition, but in terms of vertical tongue position, the accented /e/ is lower. This is consistent with the sonority expansion hypothesis.

The tongue contours for the consonantal nucleus show that the tongue tip constriction is already present before the release of the /p/, but there is no distinction in tongue tip position for the two accent conditions. From the current representation of the tongue contours it is not possible to tell whether there is actually a strong tongue tip constriction. A previous experiment on Slovak found that /l/ in nuclear position retains the tongue tip gesture (Pouplier and Beňuš, 2011), so we expect it to be the case here as well. A slightly more retracted tongue back and slightly lower tongue body when accented is again in agreement with the sonority expansion hypothesis, but also with hyperarticulation, since for the vocalic gesture they go hand in hand. In sum, there is evidence for hyperarticulation in both gestures of /l/, even for the tongue tip constriction in which hyperarticulation goes against sonority expansion.

Our data show that in principle consonantal constrictions in nucleus position are able to carry prosodic structure.

References

Beckman, M. E., Edwards, J., and Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. *Papers in laboratory phonology II*, pages 68–86.

De Jong, K. J. (1995). The supraglottal articulation of prominence in english: Linguistic stress as localized hyperarticulation. *The journal of the acoustical society of America*, 97(1):491–504.

Král', Á. (2005). *Pravidlá slovenskej výslovnosti: systematika a ortoepický slovník*. Matica slovenská.

Pouplier, M. and Beňuš, (2011). On the phonetic status of syllabic consonants: Evidence from Slovak. *Laboratory Phonology*, 2(2).

Sproat, R. and Fujimura, O. (1993). Allophonic variation in english/l/and its implications for phonetic implementation. *Journal of phonetics*, 21(3):291–311.

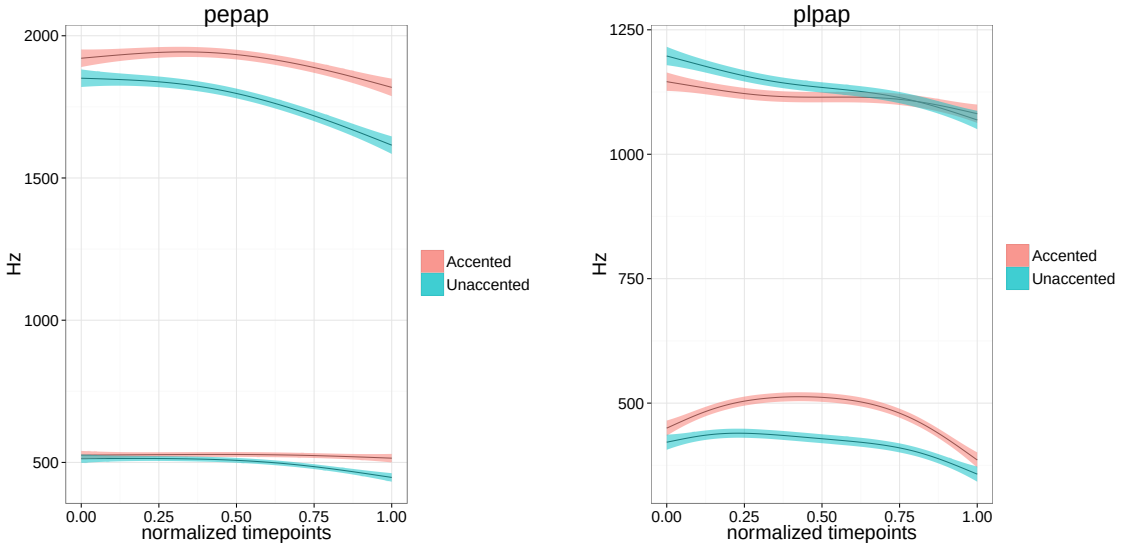


Figure 1: Smoothing Spline ANOVAs of time normalized formant movement throughout the target nucleus in *pepap* on the left and *plpap* on the right for the accented and unaccented condition

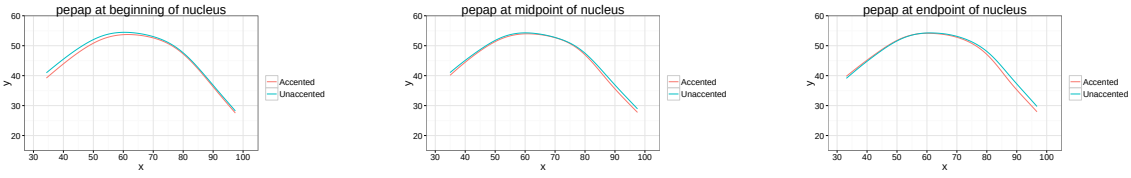


Figure 2: Mean tongue contours for pepap

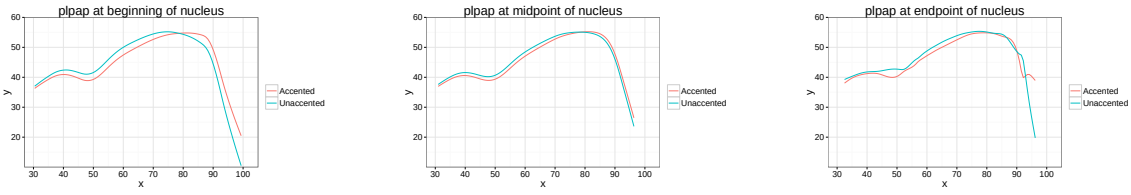


Figure 3: Mean tongue contours for plpap

GetContours: an interactive tongue surface extraction tool

Mark Tiede[1,2] and D. H. Whalen[2,1,3]

[1]Haskins Laboratories, [2]CUNY Graduate Center, [3]Yale University

Automated methods for extracting 2D tongue surface contours from sequences of ultrasound images are continuing to improve in sophistication and accuracy, ranging from Active Contour models (Kass et al. 1988) as implemented in EdgeTrak (Li et al. 2005), Deep Belief Networks as implemented in Autotrace (Fasel & Berry 2010), and Markov Random Field energy minimization as implemented in TongueTrack (Tang et al. 2012). However, a need remains for simple interactive tools that can be used to seed and propagate tracings of the tongue, and to validate these methods through comparison of automatic and manual tracings.

GetContours is a Matlab-based platform that provides straightforward click-and-drag positioning of reference points controlling a cubic spline fit to a displayed ultrasound image of the tongue surface. It supports image filtering, averaging, and contrast enhancement. Praat TextGrids (Boersma & Weenink 2015) labeled on associated audio can be imported to identify and annotate articulatory events of interest, allowing rapid selection of key frames within image sequences. While GetContours provides an implementation of the Kass et al. (1988) 'snake' algorithm for automated contour tracking, it also supports a 'plug-in' interface for applying externally available alternative algorithms seeded by the current contour.

We demonstrate GetContours through a comparison of interactive and automatic tracking of sequences of midsagittal tongue shapes produced in running speech observed simultaneously with ultrasound and electromagnetic articulometry (EMA). Results are compared with the corresponding point locations of EMA sensors attached midsagittally to the speaker's tongue.

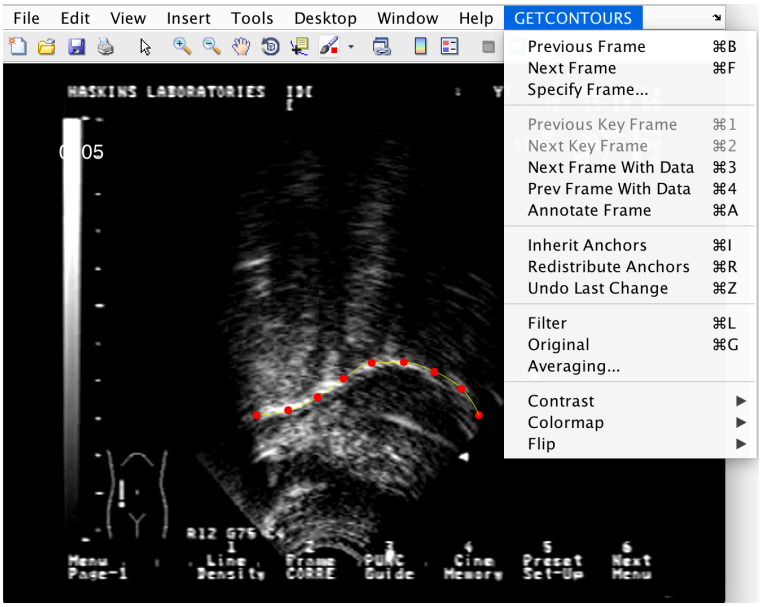


Figure 1: Illustration of US frame fit using GetContours showing available options

References

Boersma, P. & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.14, retrieved 24 July 2015 from http://www.praat.org/

Fasel, I., & Berry, J. (2010). Deep belief networks for real-time extraction of tongue contours from ultrasound during speech. In *20th International Conference on Pattern Recognition (ICPR)*, 1493-1496.

Kass, M., Witkin, A. & Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, 1(4), 321–331.

Li, M., Kambhamettu, C., & Stone, M. (2005). Automatic contour tracking in ultrasound images. *Clinical Linguistics & Phonetics*, 19(6-7), 545–554.

Tang, L., Bressmann, T., & Hamarneh, G. (2012). Tongue contour tracking in dynamic ultrasound via higher-order MRFs and efficient fusion moves. *Medical image analysis*, 16(8), 1503-1520.

The dark side of the tongue: the feasibility of ultrasound imaging in the acquisition of English dark /l/ in French learners

Hannah King & Emmanuel Ferragne
Université Paris Diderot – CLILLAC-ARP - EA 3967

Most varieties of English have traditionally been known to distinguish two allophones for the phoneme /l/: a clear variant [l] in onset position, and a dark one [ɫ], found in syllable coda. French, on the other hand, has just one allophone of the equivalent phoneme, which is largely similar to the clear variant in English. Experimental research has shed new light on the production of the English allophonic contrast. Notably, the tongue dorsum is said to retract and the tongue body to lower during the production of the dark allophone (Sproat & Fujimura, 1993). This finding conflicts with traditional generative representations of [ɫ] with the feature [+back] and with Ladefoged's analysis as velarisation (Ladefoged, 1982). As French does not have such a pronunciation and as the majority of learners in France do not undergo explicit pronunciation training prior to university, we hypothesised that French learners of English do not pronounce the dark variant in the same way as native English speakers. As the allophones of /l/ in English do not, by definition, constitute a phonemic opposition, the use of one of these allophones in all contexts would not necessarily hinder comprehension. However, if learners wish to conform to English pronunciation norms, i.e. Received Pronunciation, which is generally the variety taught in France, learning how to distinguish these two allophones is encouraged (Cruttenden, 2008). The overall aim of this study was to establish whether or not ultrasound imaging is a feasible method in a pronunciation training environment to improve French learners' acquisition of the allophones of /l/.

In order to assess this hypothesis, the tongues of 10 French learners of English and 10 native English speakers were imaged using ultrasound during the production of /l/ in various contexts (word initially and word finally, preceding and following the vowels /i/ and /u/). In order to draw comparisons between the articulations of /l/ in the two languages, French participants pronounced words in English and in French with/l/in the same context (for example, ENG "*peel*" [pi:ɫ] and FR "*pile*" [pil]).

Ultrasound data illustrated that most of our French participants do indeed distinguish the two /l/ allophones of English in their production in one way or another. It is worth noting that even amongst native Anglophone speakers, the articulation of the dark variant of /l/ varied greatly from one individual to another. This variation is almost certainly a reflection of physiological differences, as well as differences in individual pronunciation habits, and the fact that we did not control head and probe movement during experimentation, unlike other researchers have done previously (Scobbie et al., 2008; Stone, 2005). Using Edgetrak, ultrasound images were converted into a set of 30 coordinates for statistical analysis. Our data illustrated a significant difference between the average highest point of the tongue in native speakers and in French learners of English, the Anglophone tongue being in a more posterior position than the French. There was a significant difference between the light and the dark variant in both native Anglophone speakers and in learners. However, there was no significant difference between the average highest point of the dark variant for the learners and the clear for the Anglophones. We concluded that if we are able to observe differences between the tongue positions of English native speakers and those of learners during the pronunciation of [ɫ] through ultrasound visualisation, ultrasound could be a viable and effective method of direct visual feedback for learners. Other ultrasound studies have drawn similar conclusions (Gick et al., 2008; Tateishi & Winters, 2013; Tsui, 2012; Wilson, 2014).

Our next move will be to test whether the observed articulatory difference produced by French learners conveys a reliable and native-like perceptual difference. If this is not the case, then articulatory trainings with visual feedback involving ultrasound tongue imaging will be performed.

References

Cruttenden, A. (2008). *Gimson's Prononciation of English* (7th Edition). London: Hodder Education.

Gick, B., Bernhardt, B. M., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. In J. Hansen & M. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 309–322).

Ladefoged, P. (1982). *A Course in Phonetics* (2nd Edition). New York: Harcourt, Brace, Jovanich.

Scobbie, J. M., Wrench, A., & Van Der Linden, M. (2008). Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement. In *Proceedings of the 8th international seminar on speech production* (pp. 373–376).

Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics, 21(3)*, pp. 291–311.

Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics, 19*, pp. 455–502.

Tateishi, M., & Winters, S. (2013). Does ultrasound training lead to improved perception of a non-native sound contrast?: Evidence from Japanese learners of English. Presented at the Proceedings of the 2013 annual conference of the Canadian Linguistic Association.

Tsui, H. M.-L. (2012), 'Ultrasound speech training for Japanese adults learning English as a second language,' MSc Thesis, University of British Columbia.

Wilson, I. (2014). Using ultrasound for teaching and researching articulation. *Acoustical Science and Technology,* 35(6), pp. 271-289.

Searching for Closure: Seeing a Dip

Cornelia J Heyde[1], James M Scobbie[1], Ian Finlayson[1,2]

[1] *Clinical Audiology, Speech and Language (CASL) Research Centre, Queen Margaret University, Edinburgh, UK*
[2] *School of Philosophy, Psychology and Language Sciences (PPLS), Edinburgh University, Edinburgh, UK*

Quantifying lingual kinematics in relation to passive articulators is as crucial and elementary as it is challenging for ultrasound tongue imaging (UTI) research. In UTI, generally only the active tongue is observable, with passive articulatory structures such as the hard and soft palate being invisible almost all of the time. The fact that the tongue can take on various lengths and an almost indefinite set of shapes further accounts for the difficulty in establishing a referent that would allow for inter-speaker comparison. Finding a referent that respects articulatory heterogeneity is a persistent challenge. In the case of a velar stop, for example, how is the constriction found in the image? Frisch (2010) has argued for the value of automatic detection of the location of a constriction based on the shape of the tongue surface as it is deformed by contact, thereby relying on the tongue shape itself. Another approach that avoids external referents is that of Iskarous (2005) who has investigated pivot points to explore patterns in tongue contour deformation in dynamic data.

In the current study we propose a method that uses both dynamic data and movement patterns to establish the location of the constriction. The method serves to identify a referent/measurement vector along which tongue motion during the approach to and movement away from a constriction can be measured speaker-independently. We report the use of this novel technique as applied to velar closures. The resulting measures obtained along the vector can be used to quantify the degree and timing of lingual movement before and after closure, while also identifying the location of the constriction.

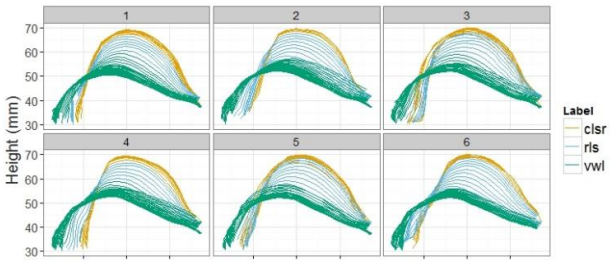


Figure 1 - splined tongue contours (tongue root on the left; tongue tip on the right) for six productions of the same /kɑ/ prompt produced by the same speaker B

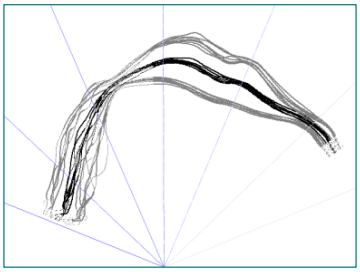


Figure 2 - overlaid mean splines (black) and SDs (grey) for the six productions of /kɑ/ (Figure 1)

The technique takes as its input multiple tokens of /kV/ targets which have been semi-automatically splined for about 700 ms (Figure 1; Articulate Instruments Ltd 2012). A fan-shaped grid of 42 equidistant radial fanlines is superimposed (Figure 2). The polar co-ordinates at which each fanline intersects with the spline are recorded. This allows us to calculate the distance to the surface from a virtual origin located within the ultrasound probe. Distances from the probe to the tongue surface at adjacent fanlines are clearly going to be highly correlated. We plotted these correlations (Pearson's r; Figure 3) for

splines that were extracted from the acoustic midpoint of the closure and found they can be used to guide the placement of a measurement vector, a fanline. As expected, there was always an extremely high correlation of the polar distances to the tongue surface of adjacent fanlines as calculated across repetitions of the same phoneme.

We noticed however a 'dip' (in a few cases we observe multiple dips, such as for speakers I and K in the bottom left and bottom right panels of Figure 3) that occurs in the midst of the overall high correlations of each speaker's correlated data. Plotting r for all adjacent fanline pairs along the tongue surface therefore results in high correlations at adjacent splines on the tongue surface, generally speaking, with a dip in correlation of two fanlines. The correlation dips stand for a reduced reliability of the location of the tongue spline for the respective area. In all but one of the cases (cf. speaker A in the upper leftmost panel) the most prominent correlation dips occur relatively central (near fanline 21) to the correlated fanlines of the ultrasound image, which is also where we would expect the tongue to form the palatal constriction in the case of /k/.

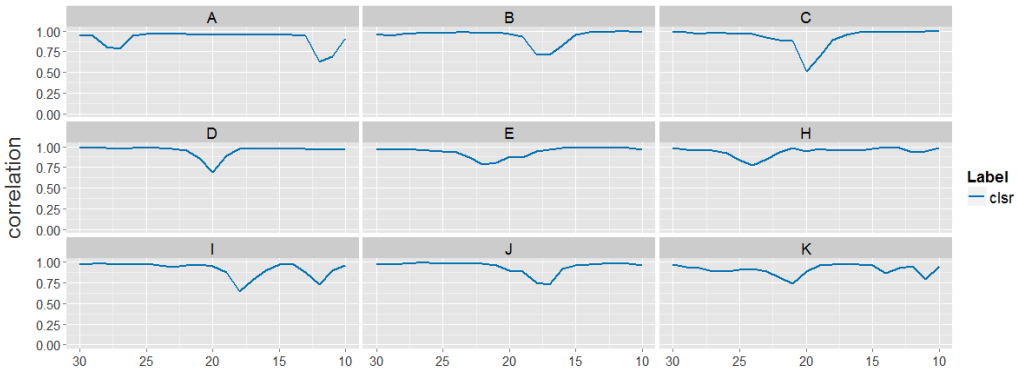


Figure 3 - Pearson's r correlations of adjacent radial fanlines along the tongue surface (from left = posterior to right = anterior) across multiple repetitions of /kV/ produced by 9 speakers (A – K)

In a previous study also on the formation of velar closure (also including the data for the current study) we have semi-manually established the fanline along which the extent of lingual movement is greatest. Interestingly, we found a meaningful overlap of those semi-manually established fanlines and the fanlines marked by the correlation dip in the current study. The systematic occurrence of the dips in addition to the clear overlap of their location and that of the semi-manually found fanlines is intriguing.

Together this indicates that dips are more than random occurrences. Dips are likely to be related to the closing gesture from which the splines were extracted. They may be particularly useful in the study of motor control as they may indicate: (1) the location of the tongue at closure and/or (2) the accuracy with which the tongue moves into the closing gesture.

A particularly interesting potential interpretation is that the dips occur where the part of the tongue that is bent behaves most circumferential to the fanlines. Any variation in the tongue contour at the point of constriction is likely to be equidistant from the probe merely shifting perpendicularly to the fanlines rather than varying in distance from the probe. The circumferential shifting along the fanline results in increased variability in that particular area because the tongue contour will be crossing the particular fanline at different slopes for each recording. At the time of consonantal closure, the most convex and also most circumferential part of the tongue is the part that touches the palate. Dips therefore capture the noise in the data that stems from the fact that over multiple

repetitions the tongue varies perpendicularly to the fan, with the variation of the most circumferential part (at the most arched part of the tongue) causing the dip.

In our interpretation, slope variation at the most convex part of the tongue contour is the cause for decreased correlation values in the relevant location. Dips indicate where the variation is largest, allowing placement of a vector to measure the kinematics of the stop in the relevant location. Dips may therefore be useful to obtain inform about the articulatory stability of a speaker. The location, steepness and width of the dip may serve as an indicator of how consistently closures are produced across repetitions.

This approach may provide information about coarticulatory processes. The approach is relatively speaker independent though it has its limitations as some speakers' oral cavities or articulation appear to be too far from typical such as , for example, speaker A in the top left panel of Figure 3. Further, measuring the tongue surface movement along the vector that crosses the dip (i.e., the measurement vector) can inform about displacement, velocity and duration of articulatory movement strokes from dynamic data. In contrast to attempts to establish an external referent, dips are inherent to the data, rendering external referents superfluous.

References

Articulate Instruments Ltd (2012). *Articulate Assistant Advanced User Guide: Version 2.14*. Edinburgh, UK: Articulate Instruments Ltd.

Frisch, S. A. (2010). Semi-automatic measurement of stop-consonant articulation using edgetrak. *Ultrafest V*.

Iskarous, K. (2005). Patterns of tongue movement. *Journal of Phonetics*, *33*(4), 363-381.

A thermoplastic head-probe stabilization device

Anna Matosova, Lorenzo Spreafico, Alessandro Vietti, Vincenzo Galatà

Free University of Bozen-Bolzano

When collecting ultrasound tongue images it is necessary to stabilize the ultrasound transducer along the midsagittal plane to avoid deviations in measurement data. Many methods exist for holding the probe in a stable relationship relative to the head. One of the most used technique has the transducer attached to a helmet that extends under the speaker`s chin, which is also the preferred solution for field work. Probably the most used head-probe stabilization headset is the one designed, manufactured and sold by Articulate Instruments [1, 2]. Over the years, the system has been refined and produced in different shapes and materials, including polycarbonate to allow co-registering ultrasound and electromagnetic articulometry data.
In this poster, we present the preliminary results of a research aimed at testing if the head-probe stabilization headset can still be improved. We consider the following areas of possible improvement:

Manufacturing: The production of metallic headsets made of rigid aluminum and of non-metallic headsets made of polycarbonate is cost and time consuming. Typically, head-probe stabilization helmets are made of more elements that need be cut, bent, milled, finished, glued and manually assembled. Here we propose a 3D printing procedure to make an easily assemblable three-dimensional object made of a limited number of thermoplastic components with no metallic inserts. The additive manufacturing methods we propose eases the production of curved elements. On the one side, this allows implementing a truss structure for the head-mount, thus characterized by both stiffness and lightness. On the other side, the 3D printing procedure permits molding shapes that are more anatomical. Both solutions guarantee more *comfort* to the speakers wearing the headset.

Usability: The headset set up can be lengthy and stressful for the informant as multiple adjustment are required to find the better tuning. In order to abbreviate and simplify the procedure, we propose to use buttons instead of lock screws. Buttons are installed on the probe-mount. The probe-rest is detached from the head-mount, but the two components can easily be connected to each other using linear guides. On the one side this design allows stiffening the headset. On the other side, it permits splitting the function of the two elements: only the inferior part of the headset has buttons to control the four degrees of freedom of probe adjustment. As the head-mount and the probe-rest are detached, it is possible to combine head-set that fit different head shapes with the same probe holder.

In our poster we will present advantages and disadvantages of the proposed solution, as well as its reliability for data collection, and contrast it with other solutions on the market.

Fig. 1: Sketch of the headrest

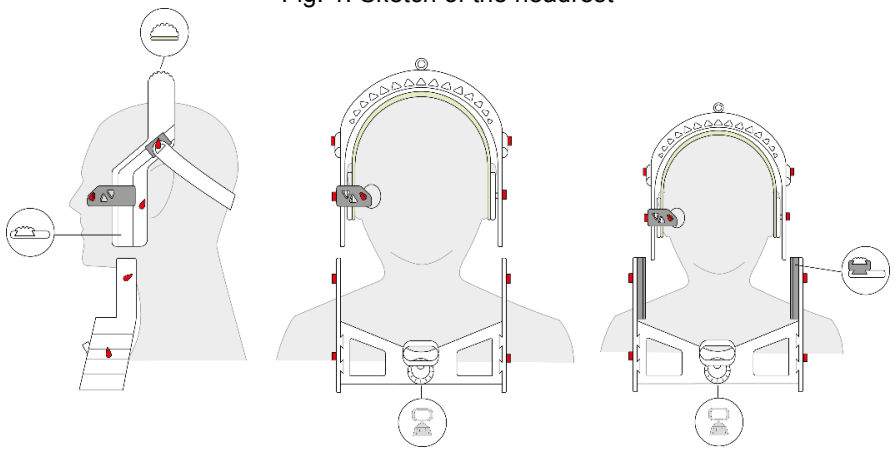
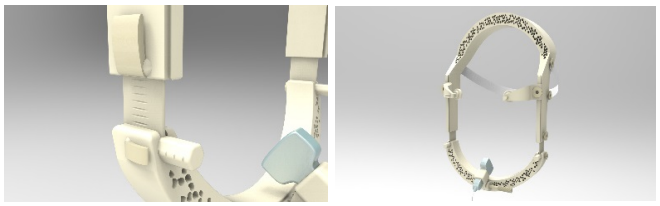


Fig. 2: Render of the headrest



[1] Scobbie, J.M., Wrench, A.A., and Linden, M. van der, (2008), *Head-Probe Stabilisation in Ultrasound Tongue Imaging Using a Headset to Permit Natural Head Movement*, Proceeding of 8th Internation Seminar on Speech Production, Strarbourg.
[2] Sigona, F., Stella, A., Gili Fivela, B., Montagna, F. Maffezzoli, A. Wrench, A, Grimaldi, M., (2013) A New Head - Probe Stabilization Device for synchronized Ultrasound and Electromagnetic Articulography recordings. Ultrafest VI, Edinburgh.

Ultrasound-Integrated Pronunciation Teaching and Learning

Noriko Yamane, Jennifer Abel, Blake Allen, Strang Burton, Misuzu Kazama, Masaki Noguchi, Asami Tsuda, and Bryan Gick

University of British Columbia

1. Introduction

Pronunciation is an integral part of communication, as it directly affects speakers' communicative competence and performance, and ultimately their self-confidence and social interaction. Second language (L2) pronunciation is one of the most challenging skills to master for adult learners. Explicit pronunciation instruction from language instructors is often unavailable due to limited class time; even when time is available, instructors often lack knowledge of effective pronunciation teaching and learning methods. Imitating native speakers' utterances can be done independently from classroom learning, but the absence of feedback makes it difficult for learners to improve their skills (e.g., de Bot, 1980; Neri et al., 2002). As well, learning to articulate difficult or unusual sounds can be made more challenging when learners have only auditory input, as the mapping from sound to articulation is not always straightforward (e.g., Wilson & Gick, 2006; Gick et al., 2008).

In an effort to improve pronunciation instruction, the Department of Linguistics and the Japanese language program in the Department of Asian Studies at the University of British Columbia began a collaboration in 2014 designed to develop new multimodal approaches to pronunciation teaching and learning. The Japanese language program is the largest language program at UBC, with more than 1,500 students enrolled every year, and is also known to be the most diverse in terms of learners' language backgrounds. The project is developing online resources to allow learners of Japanese to improve their pronunciation, as well as to allow Linguistics students to better understand sound production. The key technological innovation of this project is the use of ultrasound overlay videos, which combine mid-sagittal ultrasound images of tongue movement in speech with external profile views of a speaker's head to allow learners to visualize speech production. This technology is currently being extended to create an interactive tongue visualizer, which will allow learners to see their lingual articulations overlaid on video of their head in real time.

2. Methods

Ultrasound of native speakers of Japanese and of English was recorded using an Aloka ProSound SSD-5000 system, and the exterior video was recorded using a JVC camcorder (GZ-E300AU). Both recordings were made at 30 frames per second. The exterior video showed the left profile of the speaker's head. A clapper was used to generate an audio alignment point.

The ultrasound overlay videos were created from raw footage using a four-step process. First, the ultrasound and exterior video were trimmed using Adobe Premiere to ensure alignment. Next, all elements of the ultrasound image aside from the tongue were manually erased using Adobe After Effects. The brightness of the tongue was increased, and the colour was changed from white to a shade of pink (colour #DE8887 in Adobe After Effects) to more closely resemble the human tongue. Then, the erased ultrasound image was overlaid on the exterior face video using Adobe After Effects. Scaling of the two sources was achieved by ensuring that the shadow of the probe in the ultrasound image is the same width as the top of the probe in the exterior video. The results of this process are exemplified in Figure 1.

Figure 1. Ultrasound overlay video frame of [χ]

3. Results

The videos are available to the public through the eNunciate website (http://enunciate.arts.ubc.ca/), and are licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. The videos are categorized into 'Linguistics' and 'Japanese' content, although all pages are open to all students and instructors.

3.1 Linguistics content

The Linguistics pages feature ultrasound overlay videos of canonical examples of the sounds of the world's languages produced in basic contexts: in the [_a] and [a_a] positions for consonants, and in isolation or in [C_C] contexts for vowels. Freeze frames are inserted in these videos to capture key moments in the articulation (e.g., the stop closure in a stop articulation), and beginning and end titles are inserted. These videos can be accessed through interactive IPA consonant and vowel charts. In addition to the ultrasound overlay videos, videos introducing the use of ultrasound in linguistics and the basics of vowel and consonant articulation are available.

In the Fall 2015 term, four UBC Linguistics courses used the resources: two general introductory courses (Linguistics 100 and 101), one introduction to phonetics and phonology course (Linguistics 200), and one upper-year acoustic and instrumental phonetics course (Linguistics 313). In Linguistics 200, of the 26 students who responded to a voluntary survey, 23 (88%) indicated that the resources were easy to use and that they helped them understand how sounds are articulated, 21 (81%) indicated that the resources helped them understand the differences between sounds, and 24 (92%) indicated that they would recommend the resources to other students. Data collection on student use of and satisfaction with the resources from these courses is ongoing.

3.2 Japanese content

The Japanese pages include instructional and exercise videos for Japanese pronunciation teaching and learning. These videos incorporate narration, cartoons, and animations in addition to ultrasound overlay elements, and are augmented with quizzes to allow students to reinforce what they have learned using the videos. The videos are grouped into three categories: introductory, which includes introductions to Japanese sounds and to basic phonetic concepts; 'challenging sounds', which features videos focusing on problems that L2 learners from different language backgrounds may encounter; and intonation.

In the Fall 2015 term, the eNunciate video resources are being used in two sections of the beginner-level Japanese 102 course, which are taught by the same instructor. In one of these sections, the student will also receive a half-hour ultrasound feedback session with the first author to help improve their pronunciation. These sections are being compared with a third section, also taught by the same instructor, in which neither eNunciate resources nor ultrasound feedback are being used, to determine if use of these resources will lead to a greater improvement in students' Japanese pronunciation than 'traditional' pronunciation practice.

Table 1. Implementation of eNunciate resources and ultrasound feedback session in three sections of Japanese 102 at the University of British Columbia.

	'Traditional' Pronunciation Practice Section	**Pronunciation Practice with eNunciate Section**	**Pronunciation Practice with eNunciate and Ultrasound Section**
Activities	• Shadowing • Listening to Audio	• Watching eNunciate 'Challenging Sounds' Videos • Listening to Audio	• Watching eNunciate 'Challenging Sounds' Videos • Ultrasound Feedback Session
Assessment by Students	• Survey • Reflection Paragraph	• Survey • Reflection Paragraph	• Survey • Report on Ultrasound Feedback Session
Assessment of Students	• Perception Test • Recording Assignment	• Perception Test • Recording Assignment	• Perception Test • Recording Assignment

Data collection on student use of and satisfaction with the resources from these courses is ongoing.

4. Discussion: developments in progress

4.1 Interactive tongue visualizer

As part of our plan to use biofeedback to facilitate L2 pronunciation learning, we are developing an interactive tongue visualizer, which will automate creation of the type of ultrasound overlay videos described in section 2 based on ultrasound and video feeds of a speaker producing sounds in real time. Development of this tool is still in the early stages. The visualizer will be implemented at a physical location ("Pronunciation Station") at UBC, and will be equipped with a CHISON ECO 1 portable ultrasound with a 6.0MHz D6C12L Transvaginal probe.

4.2 Ultrasound training

To overcome the lack of a standardized procedure for the teaching of L2 pronunciation with ultrasound imaging, we are developing guidelines based on the procedures previously used in the settings of L2 learning (Gick et al., 2008) and speech language pathology (Bernhardt et al., 2005). The guidelines target three consecutive days of teaching to allow teachers to use the Pronunciation Station: (1) initial evaluation of students' pronunciation, (2) training with ultrasound images as biovisual feedback, and (3) post-training evaluation of students' pronunciation. As a case study, we implemented the protocols in teaching Japanese pronunciation to four native speakers of Korean, particularly focusing on the acquisition of the contrast between alveolar and alveo-palatal sibilants (e.g. [za] vs. [ʑa]), which is known to be especially difficult for Korean speakers. The results suggest that the protocols are effective:

the two beginner learners, one advanced learner, and one heritage speaker, none of whom had any significant contrast between those sounds in pre-training recording, showed a significant contrast in post-training recording.

4.3 Expansion to additional languages

In 2016, we intend to begin development for materials for additional languages being taught at UBC: Chinese, French, Spanish, German, and English as a second/additional language.

Acknowledgements

This project is supported by a Flexible Learning Large Project Grant from the Teaching and Learning Enhancement Fund at the University of British Columbia. Many thanks to Joe D'Aquisto, Jonathan de Vries, Amir Entezaralmahdi, Lewis Haas, Tsuyoshi Hamanaka, Hisako Hayashi, Bosung Kim, Ross King, Andrea Lau, Yoshitaka Matsubara, Douglas Pulleyblank, Nicholas Romero, Hotze Rullmann, Murray Schellenberg, Joyce Tull, Martina Wiltschko, Jenny Wong, and Kazuhiro Yonemoto.

References

Bernhardt, B., et al. (2005). Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics & Phonetics* 19.6-7: 605-617.

de Bot, C. L. J. (1980). The role of feedback and feedforward in the teaching of pronunciation. *System*, 8, 35-45.

Gick, B., et al. (2008). Ultrasound imaging applications in second language acquisition. In J. G. Hansen Edwards and M. L. Zampini (eds.), *Phonology and Second Language Acquisition* (pp. 309-322). Amsterdam: John Benjamins.

Neri, A., et al. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, 15(5), 441-467.

Pillot-Loiseau, C., et al. (2015). French /y/-/u/ contrast in Japanese learners with/without ultrasound feedback: vowels, non-words and words. Paper presented at ICPhS 2015. Retrieved August 12, 2015 from http://www.icphs2015.info/pdfs/Papers/ICPHS0485.pdf.

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

Development of coarticulation in German children: Acoustic and articulatory locus equations

Elina Rubertus[a], Dzhuma Abakarova[a], Mark Tiede[b], Jan Ries[a], Aude Noiray[a]

[a] University of Potsdam, [b] Haskins Laboratories

The present study investigates the development of coarticulation in German children between 3 and 7 years of age. To quantify coarticulation degree, we will not only apply the commonly used method of Locus Equations (LE) on the acoustic signal, but also on the articulation recorded with ultrasound, which so far has been rarely done in children (Noiray et al., 2013). This allows us to directly track dynamic movements instead of inferring (co)articulation from the acoustic signal.

Coarticulation can be viewed as connecting single speech sounds by varying degrees of articulatory overlap. While some aspects of coarticulation are claimed to be universal, resulting from anatomic properties (e.g., overlap of labial consonants and lingual vowels), others are not that predictable and may be language-specific (e.g., vowel-to-vowel coarticulation). The way children acquire the coarticulatory patterns of their native language has been discussed intensively (i.e., holistic versus segmental theory). The present study extends previous work by investigating coarticulation with a broader set of phonemes, multiple age groups, and in both acoustics and articulation.

Five cohorts of monolingual German children (3 to 7 years of age) as well as an adult control group are tested. Stimuli are elicited in a repetition task embedded in a child friendly setting. The prerecorded acoustic stimuli consist of disyllabic pseudo words following the pattern $C_1V_1C_2V_2$, preceded by the carrier word "eine" (/a͡ɪnə/). Within the stressed first syllable (C_1V_1), C_1 is /b/, /d/, /g/, or /z/ and V_1 one of the tense, long vowels /i/, /y/, /u/, /a/, /e/, and /o/. The second CV syllable consisting of the same consonant set as C_1 plus the neutral vowel /ə/ is added to the syllable of interest such that C_2 is never equal to C_1, resulting in three different contexts per C_1V_1. In total, there are 72 different pseudo words. Besides the CV coarticulation within the pseudo word, the carrier phrase enables the investigation of V-to-V anticipatory coarticulation from V_1 on the preceding schwa. At Ultrafest VII we will present the first results for CV coarticulation in the cohort of 5 year-olds and adults.

During the recordings, children are comfortably seated in an adjustable car seat. They are recorded with a portable ultrasound system (Sonosite Edge, sr: 48Hz) with a small probe fixed on a custom-made probe holder. The probe holder was designed to allow for natural vertical motion of the jaw but prevent motion in the lateral and horizontal translations. It is positioned straight below the participant's chin to record the tongue on the midsagittal plane. Ultrasound video data are collected with synchronized audio speech signal (microphone Sennheiser, sr: 48 KHz) on a computer. In addition to tongue motion, a video camera (Sony, sr: 50Hz) records the participant's face to track the labial articulation as well as head and probe motion enabling us to correct the data from a jaw-based to a head-based coordinate system.

As for the analysis, target words in the acoustic speech signal as well as relevant tongue data are extracted using custom-made Praat and Matlab programs. Acoustic LE measures of the CV coarticulation will be based on the F2 transitions between the very onset of V_1 and its midpoint, while the articulatory analysis will focus on the highest tongue point's motion between C_1 and V_1. As the ultrasound allows us to track motion earlier than is visible in the acoustic signal, we will not only use the onset of the vowel but move further into the consonant to find early cues of the vowel's influence on the tongue shape.

Development of coarticulation in German children: Mutual Information as a measure of coarticulation and invariance

Dzhuma Abakarova, Khalil Iskarous, Elina Rubertus, Jan Ries, Mark Tiede, Aude Noiray

The study aims to investigate the development of coarticulation in 3- to 7-year old German children. At Ultrafest, we present the results for 5-year-olds and adults. We try to characterize the maturation of speech motor system by looking into the way different aspects of consonant production vary on quantitative coarticulation/invariance scale as a function of age. Mutual Information (MI), a method that has been used to measure coarticulation degree by quantifying independence between two variables in adults (Iskarous et al., 2013) is adapted to the developmental field. For coarticulation, it measures the amount of information about segment B that is present during the production of segment A. MI between contiguous segments is large under coarticulation and small if the segments are relatively independent. For each consonant, we can determine the degree of independence for each of its articulators (e.g. various points on the tongue, lips, jaw). Thus, the MI method allows us to generalize the results obtained with other methods that rely heavily on tongue motion (e.g. LE) to more articulators.

Four cohorts of monolingual German children (3 to 7 years of age) as well as an adult control group are tested at LOLA Lab (Germany). Stimuli are elicited in a repetition task embedded in a child friendly setting. The prerecorded acoustic stimuli consist of disyllabic $C_1V_1C_2V_2$ pseudo words preceded by the carrier word “eine” (/a͡ɪnə/). Within the stressed first syllable (C_1V_1), C_1 is /b/, /d/, /g/, or /z/ and V_1 one of the tense vowels /i/, /y/, /u/, /a/, /e/, and /o/. The second CV syllable consisting of the same consonant set as C_1 plus the neutral vowel /ə/ is added to the syllable of interest such that C_2 is never equal to C_1, resulting in three different contexts per C_1V_1. In total, there are 72 different pseudo words.

During the recordings, children are comfortably seated in an adjustable car seat. They are recorded with a portable ultrasound system (Sonosite Edge, sr: 48Hz) with a small probe fixed on a custom-made probe holder. The probe holder was designed to allow for natural vertical motion of the jaw but prevent motion in the lateral and horizontal translations. It is positioned straight below the participant’s chin to record the tongue on the midsagittal plane. Ultrasound video data are collected with synchronized audio speech signal (microphone Sennheiser, sr: 48 KHz) on a computer. In addition to tongue motion, a video camera (Sony, sr: 50Hz) records the participant’s face to track the labial articulation as well as head and probe motion enabling us to correct the data from a jaw-based to a head-based coordinate system.

Up to now, MI metrics has been only used to quantify articulatory data from from EMA corpora. In this study, we will extend the MI metrics to a different form of articulatory data quantification, i.e. ultrasound. We will also extend the set of German consonants described with respect to their position on the coarticulation/invariance scale. Last but not least, the method allows us to quantify the changes in the position of certain consonants on coarticulation/invariance as a function of age. MI analysis is less dependent on data distribution which can be of crucial importance for children data considering the difficulties of child data collection.

The articulation and acoustics of postvocalic liquids in the Volendam dialect

Etske Ooijevaar
Meertens Instituut (Amsterdam, The Netherlands)

In different varieties of Dutch, there is variation in the production of /l/ and /r/ (Mees and Collins 1982; Booij 1995). In postvocalic position, liquids may vocalize or delete (Van de Velde et al. 1997; Van Reenen and Jongkind 2000). This can lead to neutralization of contrasts between words with and words without a liquid (Plug 2010). In addition, tense mid vowels may neutralize to their lax counterpart before a liquid (Botma et al. 2012). Although there are many acoustic studies on Dutch /r/, the articulation of Dutch liquids has only been studied recently (Scobbie and Sebregts 2011; Sebregts 2015; Haverkamp 2015).

The present study shows an Ultrasound Tongue Imaging (UTI) and acoustic analysis of postvocalic liquids in the Volendam dialect. Speakers from different age and educational level read two texts. UTI recordings were analyzed visually for ArtMax (Articulatory Maxima, Lee-Kim et al. 2013) for the vowel (/e, ɪ, o, ɔ/) and the following consonant (/l, r, t, lt, rt/) to study neutralization of /e/ and /ɪ/ (or /o/ and /ɔ/) before a liquid, retraction of the Tongue Dorsum (TD) for /l/, and raising of the Tongue Tip (TT) for /l/ and /r/. An SS ANOVA (Davidson 2006) is performed to compare differences between tongue contours.

Preliminary results of two highly educated female speakers from Volendam show that there are similarities and differences between speakers of different age (RV, 22 years old; MdWV, 62 years old). Both speakers make a contrast between /e/ and /ɪ/ before a liquid (Fig. 1), but the contrast is smaller for the younger speaker. The TD is more retracted for coda /l/ than for onset /l/ (Fig. 2). However, the younger speaker makes a clearer difference between onset and coda /l/. Fig. 3 shows that for both speakers, there is no TT raising visible for coda /l/ in sentence-final position (vocalization). Coda /l/ does show TT raising in sentence-medial position, but only the younger speaker shows a clear contrast between onset and coda /l/, that is, TT is higher in onset /l/. For both speakers, TT gestures for coda /r/ are visible in both sentence-medial and sentence-final position, but there is no clear onset-coda pattern (Fig. 4). Acoustically, postvocalic /r/ is often realized as short /s/-like frication. Postvocalic /l/ is characterized by F2 lowering.

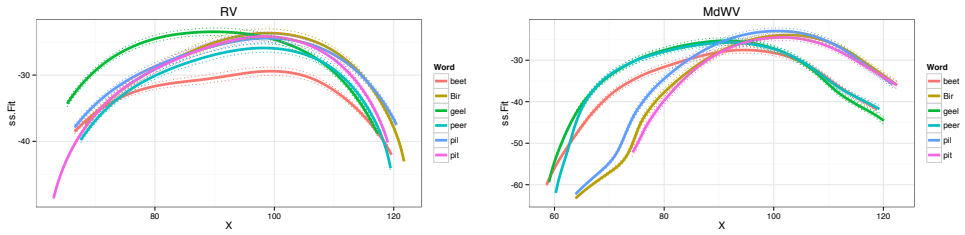


Fig. 1: ArtMax for /e/ and /ɪ/ in CVl, CVr and CVt

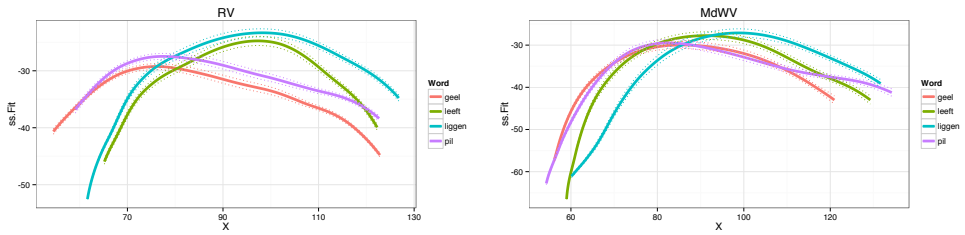


Fig. 2: ArtMax for TD in words with onset and coda /l/

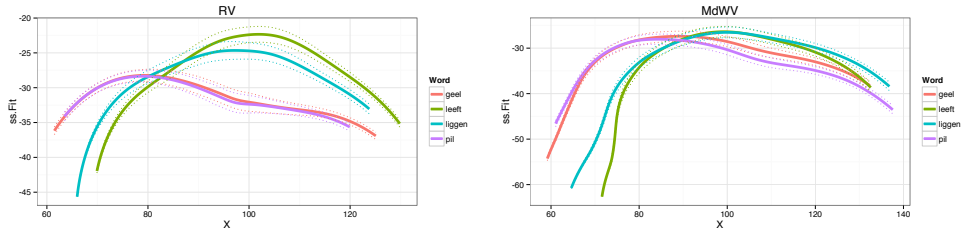


Fig. 3: ArtMax for TT in words with onset and coda /l/

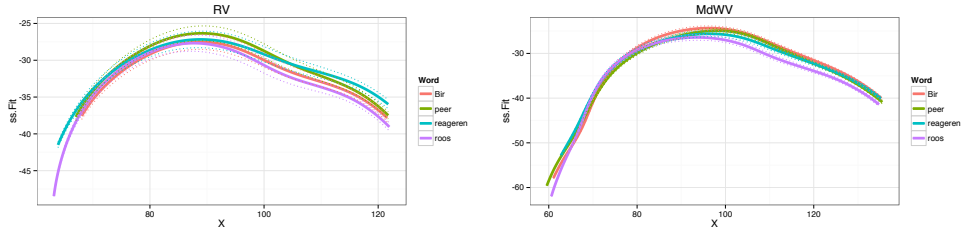


Fig. 4: ArtMax for TT in words with onset and coda /r/

Differences between MdWV and RV may show a tendency of change in articulation of liquids in Volendam. Data from more speakers will be analyzed to test whether this pattern is not just due to individual variation. In addition, the relation between articulatory data and acoustic data will also be studied.

Booij, G. 1995. The phonology of Dutch. New York: Oxford University Press.

Botma, B, Sebregts, K., Smakman, D. 2012. The phonetics and phonology of Dutch mid vowels before /l/. JLP, 3, 273-297.

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America 120,* 407-415.

Haverkamp, A.R. (2015). Palatalization as lenition of the Dutch diminutive suffix: an ultrasound study. Poster presented at CONSOLE XXIII, Paris, 7-9 January.

Lee-Kim, S.I., Davidson, L., Hwang, S. 2013. Morphological effects on the darkness of English intervocalic /l/. Laboratory Phonology, 4(2), 475-511.

Mees, I. and Collins, B. 1982. A phonetic description of the consonant system of standard Dutch (ABN). Journal of the International Phonetic Association, 12, 2-12.

Plug, L. 2010. Phonetic correlates of postvocalic /r/ in spontaneous Dutch speech. Leeds Working Papers in Linguistics and Phonetics, 15, 101-119.

Scobbie, J.M., & Sebregts, K. (2011). Acoustic, articulatory and phonological perspectives on allophonic variation of /r/ in Dutch. In: *Interfaces in Linguistics: New Research Perspectives. Oxford Studies in Theoretical Linguistics,* ed. by Folli, R., & Ulbrich, C., Oxford, OUP, 257-277.

Sebregts, K.D.C.J. (2015). *The sociophonetics and phonology of dutch r.* PhD Thesis, Utrecht University. Utrecht: Netherlands Graduate School of Linguistics LOT.

Sproat, R., Fujimura, O. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. Journal of phonetics, 21(3), 291-311.

Van de Velde, H., Van Hout, R., Gerritsen, M. 1997. Watching Dutch change: A real time study of variation and change in standard Dutch pronunciation. Journal of Sociolinguistics, 1(3), 361-391.

Van Reenen, P., Jongkind, A., 2000. De vocalisering van de /l/ in het Standaard Nederlands. In: Bennis, H.J., Ryckeboer, H., Stroop, J. (eds), De toekomst van de variatielinguistiek: Taal en Tongval 52, 189-199.

A method for automatically detecting problematic tongue traces

Gus Hahn-Powell[1], **Benjamin Martin**[1], and **Diana Archangeli**[1,2]

[1]*Department of Linguistics, University of Arizona*
[2]*Department of Linguistics, University of Hong Kong*

While ultrasound provides a remarkable tool for tracking the tongue's movements during speech, it has yet to emerge as the powerful research tool it could be. A major roadblock is that the means of appropriately labeling images is a laborious, time-intensive undertaking. In work reported at ICPR in 2010, Fasel and Berry (2010) introduced a "translational" deep belief network (tDBN) approach to automated labeling of ultrasound images. The current work extends that methodology with a modification of the training procedure to reduce reported errors (Sung and Archangeli, 2013) along the anterior and root edges of the tongue by altering the network's loss function and incorporating ℓ^1 and ℓ^2 regularization (Ng, 2004) to avoid overfitting. This training-internal approach to error reduction is compared to an independent post-processing procedure which uses the expected average positional change between adjacent points in three tongue regions (Davidson, 2006) to detect and constrain erroneous coordinates. Positional variance was calculated using the 800 most diverse and 50 least diverse tongue configurations by image pixel intensity across multiple subjects from a recitation of the phonetically balanced Harvard sentences (Rothauser et al., 1969).
Index Terms: articulatory phonetics, ultrasound imaging, tongue imaging, speech processing, deep belief networks, regularization, computer vision

References

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of Acoustical Society of America*, 120:407–415.

Fasel, I. and Berry, J. (2010). Deep belief networks for real-time extraction of tongue contours from ultrasound during speech. In *Proceedings of the 20th International Conference on Pattern Recognition*, pages 1493–1496.

Ng, A. Y. (2004). Feature selection, l 1 vs. l 2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on Machine learning*, page 78. ACM.

Rothauser, E., Chapman, W., Guttman, N., Nordby, K., Silbiger, H., Urbanek, G., and Weinstock, M. (1969). Ieee recommended practice for speech quality measurements. *IEEE Trans. Audio Electroacoust*, 17(3):225–246.

Sung, J.-H., B. J. C. M. H.-P. G. and Archangeli, D. (2013). "testing autotrace: A machine-learning approach to automated tongue contour data extraction". Edinburgh. UltraFest VI.

Word-final /r/ and word-initial glottalization in English-accented German: a work in progress

Maria Paola Bissiri and Jim Scobbie
CASL Research Centre, Queen Margaret University, Edinburgh, Scotland, UK
{MBissiri,JScobbie}@qmu.ac.uk

In Standard Southern British English, word-final /r/ is normally not articulated, as in *cider* /ˈsaɪdə/. However, /r/ can occur in connected speech if the following word starts with a vowel [1], as in *cider apples* /ˈsaɪdər ˈæpəlz/. In German, an abrupt glottalized onset to phonation is frequent in front of word-initial vowels [2], e.g. *jeder Abend (every evening)* /ˈjeːdɐ ˈʔaːbənt/, in English this is less frequent and more likely to occur at phrase boundaries and before pitch-accented words [3]. The interplay between external sandhi and glottalization is not clear: glottalizations are supposed to take place in absence of external sandhi, but articulatory gestures related to both phenomena can co-occur in a similar phenomenon, with word-final /l/ [4]. Previous investigations have shown that glottalizations are transferred in language learning [5], while the transfer of external sandhi from native to second language speech has been seldom investigated and with conflicting results [6].

We present the method and development of an ongoing study on /r/-sandhi and glottalization in English-accented German compared to English and German. By means of ultrasound tongue imaging we investigate word-final /r/ followed by a word-initial vowel, and the occurrence of glottalizations in the acoustic signal at the resulting word boundary. Accent and phrasing is also varied in the speech material.

In the present study, native English and native German speakers read sentences in both languages. Each sentence contains two subsequent words with W_1 ending with /r/, /n/ or a high vowel, and W_2 starting with a low vowel. Sentences are constructed with and without a phrase boundary between W_1 and W_2, and with W_2 accented and deaccented, thus producing four possible sentence types. We formulate the following hypotheses:

1. In the English speakers' productions, glottalizations are most frequent in the accented post-boundary condition, and sandhi is most frequent in the deaccented phrase-medial condition.
2. Sandhi is blocked by phrase boundaries, not by glottalizations, overlap between glottalization and sandhi can occur in phrase-medial position.
3. English natives transfer the extent and nature of external sandhi and glottalization in their native language to their German productions.

[1] Cruttenden, A. and Gimson, A.C. 1994. Gimson's pronunciation of English (fifth edition), revised by Alan Cruttenden. London: Edward Arnold.

[2] Kohler, Klaus. 1994. Glottal stops and glottalization in German. Data and theory of connected speech processes. Phonetica 51. 38-51.

[3] Dilley, L., Shattuck-Hufnagel, S. and Ostendorf, M. 1996. Glottalization of word-initial vowels as a function of prosodic structure. Journal

[4] Scobbie, J., Pouplier, M. (2010). The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction in word-final English /l/. Journal of Phonetics 38, 240-59.

[5] Bissiri, M.P. 2013. Glottalizations in German-accented English in relationship to phrase boundaries. In: Mehnert, D., Kordon, U., Wolff, M. (eds.), Systemtheorie Signalverarbeitung Sprachtechnologie, Rüdiger Hoffmann zum 65. Geburtstag, pp. 234-240.

[6] Zsiga, E.C., 2011. External Sandhi in a Second Language: The Phonetics and Phonology of Obstruent Nasalization in Korean-Accented English. Language, 87(2), 289-345.

The Production of English liquids by native Mandarin speakers

Chen Shuwen, Ren Xinran, Richard Gananathan, Zhu Yanjiao, Sang-Im Kim , Peggy Mok
Chinese University of Hong Kong

English liquids /l/ and /ɾ/ often present challenges to non-native speakers. In Hong Kong English, for example, the liquids are often deleted (e.g. pro[bə]m for 'problem'), replaced (e.g. [l]ide for 'ride'), or vocalized (e.g. wi[u] for 'will'). The difficulty arises partly because there is only one liquid /l/ in the inventory of Cantonese, while there are two liquids /l/ and /r/ in English. While English and Mandarin show a rough one-to-one correspondence in liquids, there are still large differences in the phonetic details of the attested liquids. For example, Mandarin speakers often vocalize the final liquid in English (Deterding, 2006) and their /l/ is notably lighter than that of American English (Smith, 2010). This indicates that the acquisition of non-native sounds is not only conditioned by the sound inventories of the first and second languages, but it is also influenced by specific distribution and phonetic details of the sounds.

Other than some descriptive studies based on subjective transcriptions, however, there is no extensive experimental data on the production of English liquids produced by Mandarin speakers. The current project aims to examine articulatory patterns in both native and non-native liquid production using ultrasound imaging. Specifically, the goals of the current study are to explore the effect of native phonological systems on production patterns and investigate detailed articulatory characteristics of foreign categories.

In the ultrasound imaging, three Mandarin speakers produced liquid sounds in Mandarin and English in three vowel contexts /ɑ i u/. For Mandarin, /ɻ/ appeared in both onset and final positions, while /l/ was limited only to initial position. The target words were embedded in short pseudo-address phrases consisting of a name of a city followed by a name of a street for Mandarin (e.g. *Menggu **L**uban Men* 'The Luban Gate in Menggu' for initial /l/ in the /u/ vowel context) and a two-digit number followed by a name of a street for English (e.g. 22 **L**oop Peak). The word lists were randomized within a language type and repeated 5 times. For comparison with native English liquids, one English speaker was recorded reading the English stimuli list.

To capture the most prototypical articulation of each liquid, the frame containing the most raised tongue front was chosen for the /ɻ/ sound and the frame containing the most retracted tongue back was chosen for the /l/ sound for both language types. The articulation of the liquids was compared using a smoothing spline ANOVA (SS ANOVA, Davidson, 2006; Wahba, 1990). Our preliminary results showed that Mandarin speakers implemented two distinct gestures for English /l/s depending on syllable position. As shown in Figure 1 (top), the initial /l/ (light grey) shows a significantly more fronted tongue dorsum than the final /l/ (dark grey). In addition, the initial /l/ appears to make alveolar contact as indicated by a significantly raised tongue blade, while such raising was not observed for the final /l/. This is suggestive of l-vocalization, but more data is needed to draw conclusions. Figure 1 (bottom) illustrates non-native /ɻ/s in initial (light grey) and final (dark grey) positions. In this particular case, a bunched /ɻ/ gesture was implemented in both positions. Full quantitative and qualitative analyses will be carried out and the results will be discussed with respect to various linguistics factors, i.e. native vs. non-native liquids, vowel effects, and positional effects.

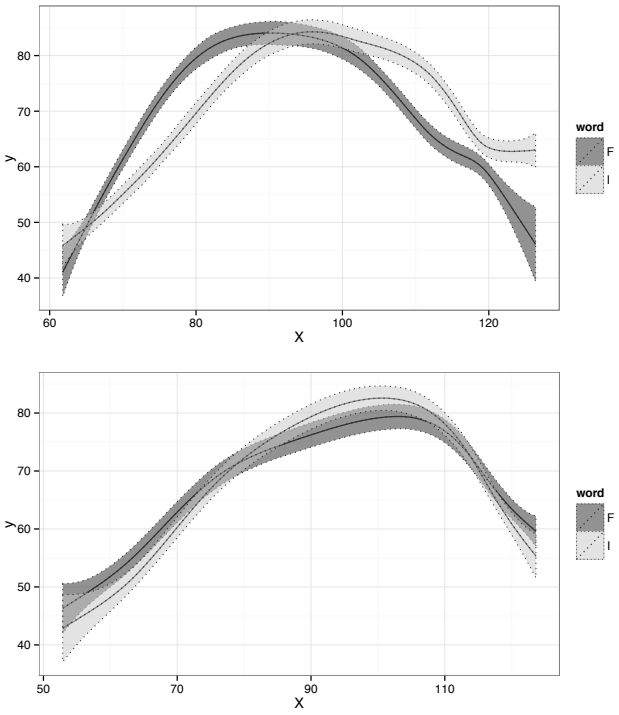


Figure 1. Smoothing spline estimate and 95% Bayesian confidence interval for comparison of the mean curves for one Mandarin speaker. **Top**. The tongue shape for the initial /l/ in *30 **L**ee Mount* (light grey) and final /l/ in *19 Pee**l** Peak* (dark grey) **Bottom**. The tongue shape for the initial /ɹ/ in *60 **R**eam Boulevard* (light grey) and final /l/ in *16 Bee**r** Peak* (dark grey). The tongue tip is on the right and the tongue dorsum is on the left.

References

Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America* 120 (1): 407-415.

Deterding, D., Wong, J. & Kirkpatrick, A. 2008. The pronunciation of Hong Kong English. *English World-Wide* 29:148–175.

Smith, J. G. 2010. Acoustic Properties of English /l/ and /ɹ/ Produced by Mandarin Chinese Speakers, University of Toronto. MA thesis, University of Toronto.

Wahba, Grace. 1990. *Spline Models for Observational Data*. Philadephia: Society of Industrial and Applied Mathematics.

Examining tongue tip gestures with ultrasound: a literature review

John M Culnan, *M.A. Program in Linguistics, HKU*

The tongue tip and blade are notoriously difficult regions to image with current ultrasound techniques. This is due to both the shadow cast by the jaw as well as the occurrence of pockets of air beneath the tip of the tongue that cause the ultrasound to reflect back before reaching the surface of the tongue (Stone 2005). The two reasons for difficulty imaging the tongue tip each carry with them unique challenges for researchers using ultrasound; this review, however, will focus only on the former. While electromagnetic midsagittal articulography (EMA) has been utilized as one alternative or supplement to ultrasound in studies where tongue tip movement is of central interest (Kochetov et al 2014, Marin & Pouplier 2013), it is not always the ideal methodology as it demonstrates the trajectory of only specific points on the tongue over time.

In the event that the tongue tip extends beyond the range of the ultrasound or is obscured by the jaw shadow, measurements may be made up to the most anterior point of the tongue that is visible (see ultrasound images in Lin et al 2014, and Miller & Finch 2011) or indicated by an additional marker on the ultrasound images, as was the case in Campbell et al. (2010). While both methods provide references relative to the image that may not correspond to the same points on the tongue, these different methods provide distinctive data, and results gathered may therefore appear divergent. Mielke and colleagues, on the other hand, used video to complete the tongue contour in their study of a Kagayanen interdental approximate (2011), which provides more accurate information.

The present literature review examines recent studies involving tongue tip gestures and evaluates the methods used for data analysis in order to bring about a discussion as to which is most effective at providing an accurate picture of the tongue across conditions, what differences, if any, may result in significant alterations of the data collected. As a final step, I recommend a simple experiment to compare these methods in order to further demonstrate the effects of this choice upon data collected.

References

Campbell, F., Gick, B. Wilson, I., and Vatikiotis-Bateson, E. 2010. Spatial and temporal properties of gestures in North American English /r/. *Language and Speech,* 53(1): 49-59.

Kochetov, A., Sreedevi, N.., Kasim, M., and Manjula, R. 2014. Spactial and dynamic aspects of retroflex production: An ultrasound and EMA study of Kannada geminate stops. *Journal of Phonetics* 46: 168-184.

Lin, S., Beddor, P., and Coetzee, A. Gestural reduction, lexical frequency, and sound change: A study of post-vocalic /l/. *Laboratory Phonology* 5(1): 9-36.

Mielke, J., Oslon, K., Baker, A. and Archangeli, D. 2011. Articulation of the Kagayanen interdental approximant: An ultrasound study. *Journal of Phonetics* 39: 403-412.

Miller, A., and Finch, B. 2011. Corrected high-frame rate anchored ultrasound with software alignment. *Journal of Speech, Language, and Hearing Research* 54: 471-486.

Stone, M. 2005. A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics,* 19(6/7): 455-501.

Index

Abakarova, D, 78, 79
Abel, J, 74
Abolghasemi, V, 10
Agostini, T, 23
Ahn, S, 22
Alexander, K, 57
Allen, B, 74
Archangeli, D, 14, 52, 82

Balch, P, 2
Beare, R, 47
Bellavance-Courtemance, M, 48
Belmont, A, 16
Benus, S, 62
Bertini, C, 49
Bissiri, M, 83
Bucar-Shigemori, LS, 62

Celata, C, 49
Chen, C, 49
Chen, S W, 84
Cleland, J, 55, 57
Coto, R, 52
Culnan, J, 86

Dawson, K, 39, 40

Erickson, D, 4

Falahati, R, 10, 49
Ferragne, E, 67
Finlayson, I, 69
Frisch, S, 16

Galatà, V, 34, 59, 72
Gananathan, R, 84
Gick, B, 74

Hahn-Powell, G, 82
Harvey, M, 23
Heyde, C, 55, 69
Howson, P, 36

Iguro, Y, 4, 54
Iskarous, K, 79

Isles, J, 57

Johnston, S, 52

Kazama, M, 74
Kim, S-I, 84
King, H, 67

Lawson, E, 17

Ménard, L, 48
Mailhammer, R, 23
Martin, B, 82
Matosova, A, 72
Maxfield, N, 16
Miller, A, 25
Mok, P, 84

Noguchi, M, 74
Noiray, A, 8, 78, 79

Ohkubo, M, 45
Ooijevaar,E, 80

Palo, P, 27
Pini, A, 59
Pouplier, M, 62

Recasens, D, 41
Reddick, K, 16
Ren, X R, 84
Ricci, I, 49
Ries, J, 8, 78, 79
Rodríguez, C, 41
Roon, K, 39, 40
Roxburgh, Z, 55
Rubertus, E, 78, 79

Schaeffler, S, 27
Scobbie, J, 17, 45, 55, 57, 69, 83
Sebregts, K, 32
Shaw, J, 23
Spreafico, L, 34, 59, 72
Story, B, 13
Strang, B, 74
Strycharczuk, P, 32

Poster
Stuart-Smith, J, 17

Tabain, M, 47
Tiede, M, 8, 39, 40, 65, 78, 79
To, C. K. S., 14
Trudeau-Fisette, P, 48
Tsuda, A, 74
Turgeon, C, 48

Vantini, S, 59
Vietti, A, 34, 59, 72
Villegas, J, 4, 54

Whalen, D, 39, 40, 65
Wilson, I, 4, 54
Wong, P, 31
Wrench, A, 2

Yamane, N, 36, 74
Yip, J, 14, 20

Zhu, Y Z, 84