Ultrafest VI (2013): Programme and Abstract booklet

9 authors, including:

James Scobbie
Queen Margaret University
111 PUBLICATIONS   **998** CITATIONS

SEE PROFILE

Claire Timmins
University of Strathclyde
23 PUBLICATIONS   **262** CITATIONS

SEE PROFILE

Zoe Roxburgh
Queen Margaret University
6 PUBLICATIONS   **3** CITATIONS

SEE PROFILE

Natalia Zharkova
Queen Margaret University
26 PUBLICATIONS   **182** CITATIONS

SEE PROFILE

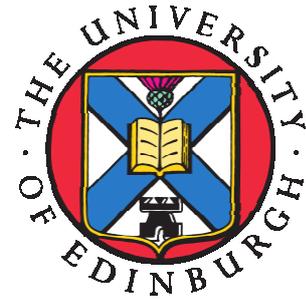**Some of the authors of this publication are also working on these related projects:**

Laboratory (socio)Phonology View project

declarative phonology View project

Programme and Abstract booklet

6th – 8th November 2013

Ultrafest VI

Welcome to the 6th Ultrafest meeting. We are pleased to welcome you all to Edinburgh.

The venue for the meeting is the University of Edinburgh Informatics Forum.

Lunch, coffee and teas will be served each day in the Atrium outside the seminar room.

The drinks reception on Wednesday night will be held in the Atrium next to the Exhibition stands.

There will be a three course meal and ceilidh at the Ghillie Dhu on Friday 8th November. We will gather outside the Informatics Forum at 6.30pm or you can make your way to the Ghillie Dhu for 7.00pm. There are limited places so please sign up for this at the registration desk.

(This is a 1.5km walk from the Informatics Forum, directions are at the back of this booklet.)

Organising Committee:

James M. Scobbie, Alan Wrench, Claire Timmins, Zoe Roxburgh, Natalia Zharkova, Eleanor Lawson, Sonja Schaeffler, Joanne Cleland, Korin Richmond

Conference sponsors:

CONFERENCE FLOOR PLAN

Presentation Room

Removable wall

Entrance

Registration Desk

Informatics Atrium

Lift

Kitchen

Servery

Buffet coffee/tea

Buffet coffee/tea

Removable wall

E1

E2

E3

E4

E5

leaflets/ Promo materials

E1-5 = Exhibition Stands

= 1x2.4m table

= Poster stand

= 1m diam tables

= 154 seats in G.07

= padded stools
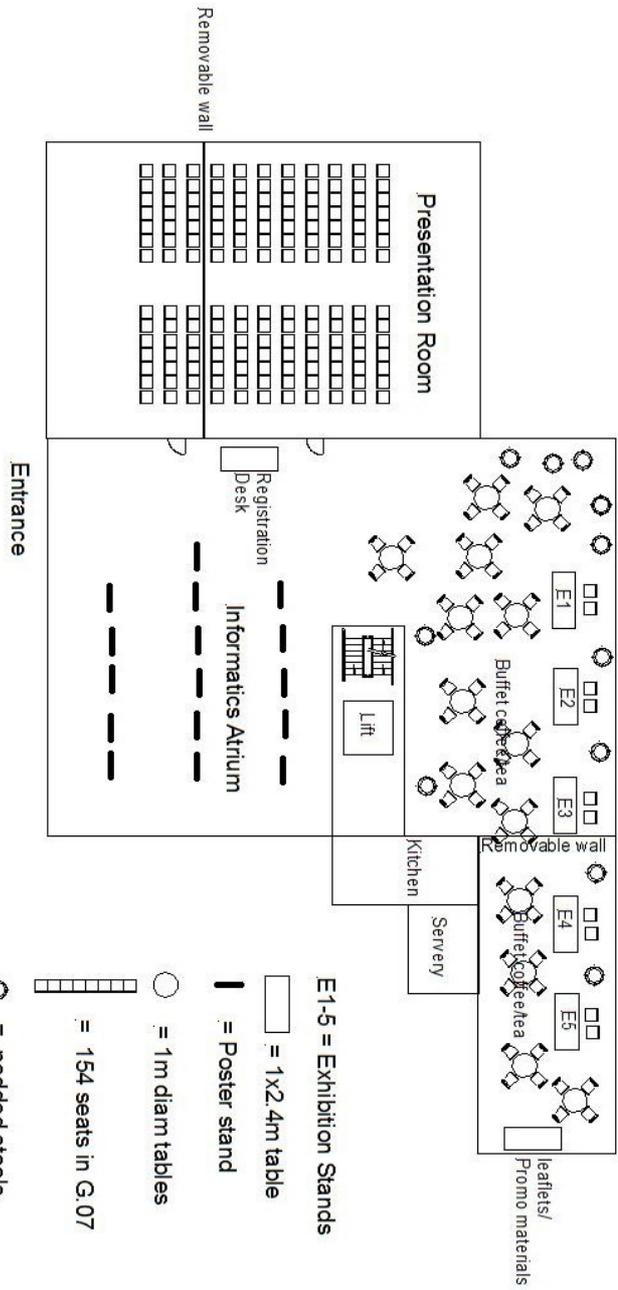
Ultrafest VI Programme

Wednesday 6th November			
8.30 – 9.15	**REGISTRATION**		
9.15 – 9.30	**Welcome:** Prof. James M Scobbie		
9.30 – 9.50	Diana Archangeli Jeff Berry Paul Boersma David Ellison Gus Hahn-Powell	**UltraPraat: Software & database for simultaneous acoustics and articulatory analysis**	Session One Chair Doug Whalen
9.50 – 10.10	Diana Archangeli Jae- Hyun Sung Marissa Cooper Gustave Hahn-Powel Jeff Berry	**Testing AutoTrace: a machine-learning approach to automated tongue contour data extraction**	
10.10 – 10.30	**COFFEE**		
10.30 – 10.50	Thomas Hueber	**Ultraspeech-tools: acquisition, processing and visualization of ultrasound speech data for phonetics and speech therapy**	Session Two Chair Doug Whalen
10.50 – 11.10	Ryan Bennett Grant McGuire Máire Ní Chiosáin Jaye Padgett	**Quantitative measures for the degree of palatalization/velarization: Irish ultrasound data**	
11.10 – 11.30	Susan Lin Benjamin Davies Katherine Demuth	**Kaytetye consonant contrasts without contours**	
11.30 – 12.30	**Discussion on Methodology and technical issues including overview:** Alan Wrench		
12.30 – 1.30 **(1.00 – 1.30)**	**LUNCH** **AAA workshop in plenary room (recording/analysis)**		
1.30 – 1.50	Eleanor Lawson James M. Scobbie Jane Stuart-Smith	**The role of gesture timing in postvocalic /r/ lenition: an ultrasound study**	Session Three Chair Diana Archangeli
1.50 – 2.10	Jeff Mielke Christopher Carignan	**Ultrasound + corpus study of phonetically-motivated North American French and English**	
2.10 – 2.30	Reza Falahati	**An ultrasound and acoustic study of consonant cluster simplification in Persian**	
2.30 – 2.45	**BREAK (Posters to be put up)**		
2.45 – 4.45	**COFFEE and POSTER SESSION 1**		
4.45 – 5.45	**KEYNOTE 1: TANJA SCHULTZ**		Chair Korin Richmond
5.45 – 7.30	**DRINKS RECEPTION**		

Thursday 7th November			
9.00 – 9.20	Sonya Bird Scott R. Moisik John Esling Peter Jacobs	Laryngeal contributions to weak vs. strong ejectives	Session Four Chair James M Scobbie
9.20 – 9.40	Sang-Im Lee Kim Lisa Davidson	Morphological effects on the darkness of English intervocalic /l/	
9.40 – 10.00	Alexei Kochetov N. Sreedevi, Midula Kasim R. Manjula	An ultrasound study of retroflex and alveolar laterals in Kannada	
10.00 – 10.30	COFFEE		
10.30 – 10.50	Marianne Pouplier Philip Hoole	Comparing principal component analysis of ultrasound images with contour analyses in a study of tongue body control during German coronals	Session Five Chair Stefan Frisch
10.50 – 11.10	Natalia Zharkova	Lingual coarticulation and articulatory constraints in adolescents	
11.10 – 11.30	Jonathan Yip	Phonetic effects on the timing of gestural coordination in Modern Greek consonant clusters	
11.30 – 11.50	Douglas H. Whalen Khalil Iskarous Lucie Ménard Amélie Prémont Katherine M. Dawson	An ultrasound analysis of tongue shape in Parkinson's Disease	
11.50 – 12.30	GENERAL DISCUSSION TIME (topic to be confirmed)		
12.30 – 1.30 (1.00 – 1.30)	LUNCH AAA workshop in plenary room (more analysis)		
1.30 – 1.50	Bryan Gick	Getting away from the midline: biomechanical devices and the role of ultrasound	Session Six Chair Natalia Zharkova
1.50 – 2.10	Lyra Maglouglin	An ultrasound study of the acquisition of North American English /ɹ/	
2.10 – 2.30	Thomas Magnuson, Cristopher Coey	What happens as listeners become speakers during face-to-face conversation?	
2.30 – 3.00	GENERAL DISCUSSION TIME (topic to be confirmed)		
3.00 – 5.00	COFFEE and POSTER SESSION 2		

Friday 8th November			
9.00 – 9.20	Amanda Miller	**Estimating lingual cavity volume in click consonant production from combined lingual ultrasound and palatographic data**	Session Seven Chair
9.20 – 9.40	Lorenzo Spreafico	**sC(C) clusters in German-Italian bilingual speakers**	Maureen Stone
9.30 – 10.00	Akitsugu Nogita Sonya Bird Noriko Yamane	**The Japanese unrounded back vowel /ɯ/ is in fact rounded central/front [ʉ - ɤ]**	
10.00 – 10.30		**COFFEE**	
10.30 – 10.50	Angelika Nair Garyth Nair Maureen Stone	**The low mandible maneuver**	Session Eight Chair
10.50 – 11.10	Ian Wilson Julian Villegas Terumasa Doi	**Lateral tongue bracing in Japanese and English**	Bryan Gick
11.10 – 11.30	Majed Al Solami	**Ultrasound and acoustic study of gutturals and emphatics in three Arabic dialects**	
11.30 – 12.30	**General discussion time (Topic: comparing tongue curves) James M. Scobbie**		
12.30 – 1.30 (1.00 – 1.30)	**LUNCH AAA workshop in plenary room (Feedback)**		
1.30 – 2.30	**KEYNOTE 2: Penelope Bacsfalvi**		Chair Tim Bressman
2.30 – 3.00		**COFFEE**	
3.00 – 3.20	Stefan A. Frisch Alissa Belmont Nathan D. Maxfield	**Velar coarticulation and tongue twisters in people who stutter 2: disfluent utterances**	Session Nine Chair
3.20 – 3.40	Tara McAllister Byun Elaine Hitchcock	**Retroflex versus bunched /r/ in ultrasound intervention: one shape does not fit all?**	
3.40 – 4.00	Tim Bressmann Irina Zhylich Gajanan V. Kulkarni	**Ultrasound biofeedback therapy for –r-distortions: is it worth the effort?**	Joanne Cleland
4.00 – 4.20	Zoe Roxburgh James M. Scobbie Joanne Cleland Sara Wood	**Can seeing help you hear? Acoustic vs. Ultrasound information in the diagnosis of speech sound disorders associated with cleft palate**	
4.20 – 4.40	**GENERAL DISCUSSION TIME (topic to be confirmed)**		
4.40 – 5.00	**CLOSING**		
7.00 – late	**DINNER AND CEILIDH (Meet outside Informatics at 6.30pm) (limited places, ask at registration desk for ticket)**		

	Poster Session One: Wednesday 6th November	
1	Aurore Jaumard-Hakoun, S. K. Al Kork, M. Adda-Decker, A. Amelot, L. Buchman, Gerard Dreyfus, T. Fux, Pierre Roussel, Claire Pillot-Loiseau, Maureen Stone, Bruce Denby	Capturing, analyzing, and transmitting intangible cultural heritage with the i-Treasures project
3	Sarah Ghrenassia, Catherine Laporte, Lucie Ménard	Statistical analysis of tongue shape in ultrasound video sequences: tongue tracking and population analysis
5	Christopher Coey, John Esling, Scott R. Moisik	Ultrasound images in the new 'iPA Phonetics' App
7	Rosey Billington	Ultrasound in language documentation: some observations on lingual movement in Lopit vowel contrasts
9	Simon Gonzalez, Mark Harvey, Susan Lin, Katherine Demuth	Segment constants and speaker constants in English coronal obstruents
11	Diana Archangeli, Daniel Brenner, Rolando Coto, Samuel Johnston	Probe skewing effects & Kinect head-tracking in tongue contour Imaging
13	Katherine M. Dawson, Douglas H. Whalen, Mark K. Tiede	A comparison of methods for tongue shape analysis
15	Natalia Zharkova, Nigel Hewlett	Quantitative analysis of ultrasound data collected with and without head stabilisation
17	Aline Mara de Oliveira Vassoler, Larissa Cristina Berti	Investigation of gestural coordination in the production of consonant clusters in children with typical development and deviant language
19	Haydée Wertzner, Danira Francisco, Luciana Pagan-Neves	Tongue contour in /ʃ/ sound for Brazilian Portuguese-speakers with SSD: a pilot study using ultrasound
21	Natalia Lapinskaya	An exploratory ultrasound investigation of emphatic articulation in Cairene Arabic
23	Laura Ryals	Glottal stop in Arabic: the insufficiency of a landmark underspecification account
25	Sarah Hamilton, Suzanne Boyce	Describing the pharyngeal constriction gesture in Turkish
27	Danielle Turton	Contextual patterns of /l/-darkening in accents of English
29	Ming Xiu, Didier Demolin	An ultrasound study of Mandarin fricatives and affricate

Poster Session Two: Thursday 7th November		
2	Stefan A. Frisch, Alissa Belmont, Nathan D. Maxfield	Velar coarticulation and tongue twisters in people who stutter 1: fluent utterances
4	Sina Koch, Tim Bressmann	An investigation of tongue movement in patients with lingual hemiparalysis
6	Jae-Hyun Sung, Diana Archangeli, Ian Clayton, Daniel Brenner, Samuel Johnston, Michael Hammond, Andrew Carnie	The articulation of Scottish Gaelic plain and palatalized consonants
8	Giovana Ferreira-Gonçalves, Mirian Rose Brum-de-Paula, Giulian Silva-Pinto	An ultrasound analysis of the acquisition of rounded French vowels /y/ and /ø/ by native speakers of Brazilian Portuguese
10	Lilian Kuhn	An ultrasound study of oral and nasal central vowels in Brazilian Portuguese
12	Tanja Kocjančič Antolík, Claire Pillot-Loiseau, Takeki Kamiyama, Cécile Fougeron	Comparing the effect of pronunciation and ultrasound trainings to pronunciation training only for the improvement of the production of the French /y/-/u/ contrast by four Japanese learners of French
14	Ai Mizoguchi, Douglas H. Whalen	An ultrasound analysis of tongue shape for the Japanese nasal mora
16	Will Dalton, Jeff Mielke	Many-to-one articulatory-acoustic mappings in Canadian French vowels
18	Jae Yung Song	Factors affecting the articulatory and acoustic characteristics of vowels in clear speech
20	Francesco Sigona, Antonio Stella, Barbara Gili Fivela, Francesco Montagna, Alfonso Maffezoli, Alan Wrench, Mirko Grimaldi	A new head-probe stabilization device for synchronized ultrasound and electromagnetic articulography recordings
22	Alessandro Vietti, Lorenzo Spreafico, Vittorio Anselmi	An ANN approach to the recognition of UTI-based Italian words
24	Sonja Schaeffler, James M. Scobbie	Enhancing Measures of Reaction Time with UTI Data
26	Korin Richmond, Steve Renals, Joanne Cleland, Jim Scobbie, Alan Wrench	Ultrax: vocal tract imaging for speech therapy and research
28	James M. Scobbie, Anna Sfakianaki	A single speaker study of Modern Greek: The effect of stress, syllable position and consonantal context on vowels, and of vowel context on intervocalic consonants
30	Kevin Roon, Eric Jackson, Hosung Nam, Mark Tiede, Douglas H. Whalen	Assessment of head reference placement methods for optical head-movement correction of ultrasound imaging in speech production

Prof. Dr.-Ing. Tanja Schultz

Cognitive Systems Lab, Karlsruhe Institute of Technology (http://csl.anthropomatik.kit.edu/)

Silent Speech Interfaces

At the Cognitive Systems Lab (CSL) we explore human-centered cognitive systems to advance human-machine interaction as well as machine-mediated human communication. We aim to benefit from the strength of machines by departing from just mimicking the human way of communication. Rather we investigate the full range of biosignals emitted from the human body, such as electrical biosignals like muscle and brain activity. These signals can be directly measured and interpreted by machines, leveraging emerging wearable, small and wireless sensor technologies. These biosignals offer an inside perspective on human physical and mental activities, intentions, and needs and thus complement the traditional way of observing humans from the outside.

In my talk I will discuss ongoing research at CSL in the area of "Silent Speech Interfaces" that rely on articulatory muscle movement. The technology is based on ElectroMyoGraphy (EMG), i.e. the capturing and recording of electrical potentials that arise from muscle activity. Speech is produced by the contraction of muscles that move our articulatory apparatus. The electrical potentials which are generated by this muscular activity are captured by surface electrodes attached to the speaker's face. The analysis and processing of these signals by machine learning algorithms allow to reconstruct the corresponding movement of the articulatory muscles and to deduct what has been said. The automatically recognized speech is output as text or synthesized as an acoustic signal. Since EMG records muscle activity rather than acoustic signals, speech can be recognized even if it is uttered silently, without any sound production. It enables several applications, such as (1) *Silent Speech Communication:* Not generating any audible signal is desirable to avoid noise pollution in quiet areas, to safeguard the privacy of information being transmitted via voice and to avoid disturbing bystanders when making telephone calls in public places, (2) *Speech Rehabilitation:* Audible spoken communication is not available for those who have lost the ability to speak due to weakness, injury or disease. EMG-augmented devices may give voice to those who are mute, (3) *Robust Speech Recognition in adverse noise conditions:* Unlike the acoustic signal, EMG signals are not vulnerable to environmental noise and adverse conditions, which seriously degrade the performance of traditional speech recognition based on acoustic signals, and (4) *Complementing Acoustic Signals:* EMG signals could improve performance of traditional speech processing systems in all conditions by bringing redundancy to the emitted acoustic wave. A brief outlook on CSL's research towards interfaces that use brain activity measured by Electroencephalography and Near Infrared Spectroscopy in order to tailor the behavior of spoken dialog systems to the users' mental workload will conclude the talk.

Please refer to http://csl.anthropomatik.kit.edu/english/silentspeechinterfaces.php for details

Dr Penelope Bacsfalvi

School of Audiology and Speech Science, Univeristy of British Columbia

Ultrasound – an essential part of the toolkit

This presentation will discuss why ultrasound is an effective tool for use by clinicians in the remediation of speech sound disorders. I will present one case to help you understand the significance of the use of ultrasound as a tool. Another focus will be on what I have learned over the last 14 years of intervention, illustrated through video and the eyes of my clients. I will also discuss the strengths and weaknesses of ultrasound with input from a client.

UltraPraat: Software & database for simultaneous acoustic and articulatory analysis

Diana Archangeli^*, Mohsen Mahdavi*, David Ellison#, Gus Hahn-Powell*, Rolando Coto*, Jeff Berry@, Paul Boersma%

*^University of Hong Kong, *University of Arizona, #VoiceBox, @InsideSales.com, %University of Amsterdam*

Currently there is no reliable open-source integrated software for extraction and analysis of articulatory data. For example, a recent study of multimodal articulatory data (Grimaldi et al. 2008) describes a complex software pipeline involving the software tools EdgeTrak, Mat- lab (both closed source), Praat and Mplayer (both open source), three scripting languages, command-line interaction, and a custom GUI. Such "baling-wire and duct-tape" solutions are untenable and a major bottleneck for progress in the field.

Rather than creating a new stand-alone tool, we are developing software to add artic- ulatory data analysis modules to enhance Praat (Boersma & Weenink 2009), a modular, open-source speech analysis program which is already the most widely used software for analysis of audio speech data (estimated to have 17,000 users worldwide). Praat currently allows researchers to perform a wide variety of analyses and manipulations to audio (mono or stereo) data, including spectral analyses, pitch analyses, formant analyses, and intensity analyses, manipulation of pitch, duration, intensity, and formants, and annotation on a time- line of audio data. Users can also automate analyses through the use of a built-in scripting language. These tools have been used to answer questions about a multitude of topics in phonology and phonetics, including language variation due to social factors, effects of (near) neutralization, and co-articulation. Furthermore, the applications of Praat have extended well beyond the domain of direct language research; UltraPraat – Praat with a component for articulatory analysis – has the potential for a similarly transformative impact.

Coupled with this effort, we are developing a database structure for acoustic and articu- latory data and accompanying metadata covering subject and study information.

In this presentation, we will lay out and demonstrate the current status of UltraPraat and the database, and sketch a vision for where to go next.

Our goal is a significant amount of discussion to help us prioritize for creating a useful tool for data storage and analysis.

References

Boersma, P. & Weenink, D. (2009) Praat: doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved May 1, 2009, from http://www.praat.org/.

Grimaldi, M., Gili Fivela, B., Sigona, F., Tavella, M., Fitzpatrick, P., Craighero, L., Fadiga, L., Sandini, G. Metta, G. (2008) New technologies for simultaneous acquisition of speech articulatory data: ultrasound, 3D articulograph and electroglottograph. In C.Delogu, M. Falconi (eds.), *Atti del Convegno Internazionale LangTech*, 81–85, Rome, Fondazione Ugo Bordoni.

Testing AutoTrace: A Machine-learning Approach to Automated Tongue Contour Data Extraction

Jae-Hyun Sung[1], **Jeff Berry**[2], **Marissa Cooper**[1], **Gustave Hahn-Powell**[1], and **Diana Archangeli**[3]

Department of Linguistics, University of Arizona, USA[1] *InsideSales.com* [2]

Linguistics, University of Hong Kong, HK[3]

Articulatory imaging is important for analyzing the rules of speech and can be utilized for many purposes such as analyzing sounds in different dialects, learning a second language, and speech therapy (Archangeli and Mielke (2005), Adler-Bock et al. (2007), Gick et al. (2008), Scobbie et al. (2008)). When analyzing phonological data, researchers can implement various experimental methods concerning articulation - EMMA, MRI, Palatography, and ultrasound. Although ultrasound is inexpensive, non-toxic, and portable, it does have a significant drawback. After the data is collected, someone must then trace the tongue surface contours, which creates a bottleneck for analyzing the results. Several different approaches to this problem have been proposed (Li et al. (2005), Fasel and Berry (2010), Tang et al. (2012)), with promising results. In this paper, we analyze the performance of the Deep Neural Network approach of Fasel and Berry (2010) at scale, and announce an open source project to further develop this method, named AutoTrace.

AutoTrace automates the process of extracting the data from ultrasound images, greatly reducing the amount of time necessary for tracing images, as shown in Berry et al. (2012) on a small data set. This paper reports on our tests of the efficacy of AutoTrace on a much larger data set consisting of approximately 40,600 ultrasound images taken from Harvard sentences read by 12 American English speakers. This ensured a wide variety of tongue shapes, due both to different speakers and to different types of sounds. For training data sets, we selected a combination of most and least diverse images based on their deviation from pixel averages, using the heuristic proposed by Berry (2012).

AutoTrace used training data sets of different sizes to learn networks. Each network was tested against the same set of 100 randomly selected images. These traces were hand-corrected by human expert tracers, and each network was retrained. The automatically traced contours were compared to traces made by human experts to gauge how well the program performed.

Four separate tests are considered:

a. Most diverse images only vs. most + least diverse images: the combination most & least gave better results.

b. n images vs. n images (ranging from 250 to 1056 images) taken at intervals of y: the larger the training set, the better the results.

c. No retraining vs. retraining: retraining produced a "41% decrease in the number of images needing hand correction" (Berry, 2012, p. 50).

d. n images from the most diverse set vs. r randomly selected images from the whole ($n = r$): the n most diverse images performed better.

Currently, the comparison of two human expert tracers shows a pixel-by-pixel average difference of 2.467 pixels per image. Comparison of machine vs. human tracers shows a pixel-by-pixel average difference of 5.656 pixels per image. We are currently examining the types of errors AutoTrace is making to see whether we can improve the training technology for better results.

References

Adler-Bock, M., Bernhardt, B. M., Gick, B., and Bacsfalvi, P. 2007. The use of ultrasound in remediation of English /r/ in two adolescents. American Journal of Speech-Language Pathology, 16(2):128–139.

Archangeli, D. and Mielke, J. 2005. Ultrasound research in linguistics. In 34th Annual Meeting of The Linguistic Association of the Southwest.

Berry, J. 2012. Machine learning methods for articulatory data. PhD thesis, The University of Arizona.

Berry, J., Fasel, I., Fadiga, L., and Archangeli, D. 2012. Training deep nets with imbalanced and unlabeled data. In Interspeech.

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.

Gick, B., Bernhardt, B. M., Bacsfalvi, P., and Wilson, I. 2008. Ultrasound imaging applications in second language acquisition. In Edwards, J. G. H. and Zampini, M. L., editors, Phonology and Second Language Acquisition, pages 309–322. John Benjamins.

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

Scobbie, J. M., Stuart-Smith, J., and Lawson, E. 2008. Looking variation and change in the mouth: developing the sociolinguistic potential of ultrasound tongue imaging. Queen Margaret University.

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.

Session Two

Ultraspeech-tools: Acquisition, processing and visualization of ultrasound speech data for phonetics and speech therapy

Thomas Hueber

CNRS/GIPSA-lab, Grenoble, France

Ultraspeech-tools is a set of software dedicated to the acquisition, the processing and the visualization of ultrasound speech data. First, we present latest developments of the acquisition tool, called *Ultraspeech* (presented at Ultrafest V). Then, we introduce two additional tools, called *Ultramat* and *Ultraspeech-player*, dedicated respectively to the processing and the visualization of ultrasound articulatory data for speech therapy and pronunciation training applications.

Ultraspeech is a standalone software allowing the synchronous and simultaneous acquisition of high-speed ultrasound, video, and audio signals, at their respective maximum temporal resolution. *Ultraspeech* has been primarily designed for a *silent speech interface*, in which ultrasound and video images of the tongue and lips, were used to drive a speech synthesizer [1]. *Ultraspeech* is built around the lightweight firewire ultrasound device *Terason T3000* and is compatible with WDM-compliant industrial cameras and ASIO sound systems. The portability of the *Terason T3000* makes *Ultraspeech* acquisition system compact and transportable. Latest developments of *Ultraspeech* focused on the re-positioning of the ultrasound probe for data acquisition in multiple sessions. We describe an embedded system for measuring the inclination of the ultrasound probe relative to the speaker's head. The proposed system is based on two sets of inertial sensors, placed respectively on the probe and on a pair of glasses. Each set of sensors is composed of one 3-axis accelerometer and one 3-axis gyrometer.

Accuracy of the proposed system is evaluated by comparing the inclination calculated by the inertial sensors with the one estimated via a standard motion capture device (such as *Optotrak*). Preliminary results show that the proposed system is relatively accurate in the context of tongue imaging.

Ultramat is a *Matlab* toolbox which provides functions for processing ultrasound data acquired using *Ultraspeech*, *i.e.* extracting visual features in the perspective of statistical analyses. It implements a standard semi-automatic edge extraction procedure which is based on anisotropic filtering and active contours (*snake*). It also implements the *EigenTongue* approach which is a statistical technique consisting in encoding an ultrasound frame by its projections onto a set of standard (tongue) configurations.

Ultraspeech-player is a standalone software dedicated to the visualization of ultrasound speech data recorded using *Ultraspeech*. *Ultraspeech*-player is designed for speech therapy protocols which use visual information to increase the articulatory awareness of a speaker. The originality of this tool resides in the fact that it combines a video retiming technique with an audio time-stretching algorithm. *Ultraspeech-player* allows the user to slow-down in real-time both the articulatory gesture and its corresponding acoustic realization (i.e. the speed of the audio signal is modified while the original pitch it preserved). This rendering technique aims at improving the way a naïve speaker perceive and *understand* a tongue gesture.

More information on *Ultraspeech-tools* can be found at www.ultraspeech.com

References

[1] Hueber, T., Benaroya, E.L., Chollet, G., Denby, B., Dreyfus, G., Stone, M. (2009) "Development of a Silent Speech Interface Driven by Ultrasound and Optical Images of the Tongue and Lips", Speech Communication, 52(4), pp. 288-300.

[2] Hueber, T., Aversano, G., Chollet, G., Denby, B., Dreyfus, G., Oussar, Y., Roussel, P., and Stone, M. (2007). "Eigentongue feature extraction for an ultrasound-based silent speech interface," Proceedings of ICASSP (Honolulu, USA), pp. 1245-1248.

Quantitative measures for the degree of palatalization/velarization: Irish ultrasound data

Ryan Bennett[2], Grant McGuire[2], Máire Ní Chiosáin[1], Jaye Padgett[2]

[1]*UCD, Dublin,* [2]*University of California, Santa Cruz*

In Irish, palatalization is contrastive for (almost) all consonants, and in all vocalic contexts:
  (1) [bɑːd] ~ [bɑːdʲ] *bád/báid* 'boat/boats'
  (2) /bʲiː/ [biː] ~ /biː/ [bˠiː] *bí/buí* 'be (imp.)/yellow'

This paper focuses on methodological issues arising from an ultrasound study of secondary articulations in the three major dialect groups of Irish (5 speakers per dialect, data recorded using a stabilized probe with an Articulate Instruments Ultrasound Stabilization Head-set). Specifically, we explore a variety of analytic methods for interpreting secondary palatalization in quantitative terms. The primary issues addressed are:
  a) How to measure the *magnitude* of tongue body movement for particular secondary articulations.
  b) How to assess the *absolute position* of the tongue body for particular secondary articulations.

Participants were asked to produce a series of target words containing word-initial consonants that varied systematically in major place, manner, and secondary articulation. Initial processing of the data involved tracing tongue surfaces with the EdgeTrak software tool (Li et al. 2004). An SSANOVA was then used to find a best-fit curve over 4-8 tracings of each segment (Gu 2002, Davidson 2006). This

method yields 'average' (fitted) curves that we take as representative of the prototypical articulation of each consonant for a particular speaker.

A challenge for the interpretation of this data is the lack of an absolute quantitative measure for the degree of tongue backing/fronting. One of our research questions is whether the articulatory realization of the palatalization contrast varies with the vocalic environment. Is the contrast in (2) a plain vs. velarized contrast in gestural terms (as suggested by the acoustics), or is it a true palatalized vs. velarized contrast, with the acoustic effects of palatalization masked by the following [iː]? Similarly, might the magnitude of the palatalization/velarization gesture vary by consonant type or context? To answer these questions, we need a static landmark that can serve as the basis for assessing the absolute horizontal position of the tongue body; finding an appropriate landmark is one of our principal methodological issues.

Anatomical landmarks (e.g. the alveolar ridge) can be established using palate tracings. However, we were unable to image the palate clearly for most of our speakers. Selecting and locating an anatomical landmark is therefore an error-prone process, especially since the shape of the palate lacks sharp discontinuities (i.e. anatomical 'points') that function as reliable landmarks. Moreover, it's not clear that such landmarks would be useful for measuring horizontal displacement (apart from the alveolar ridge, perhaps).

An alternative approach is to infer a reference point in the articulatory space based on other information about each speaker's productions. For example, the carrier phrase for the recordings contained a reference 'velarized' /k/ which we have successfully used as a relatively consistent, speaker-dependent landmark for computing the absolute position of the tongue body. It is also roughly anatomical, being at/near maximum dorsal constriction. However the absolute position of this reference /k/ is not itself known – it could be truly velarized, or more central, thus introducing a potential circularity problem. The articulatory location of the reference /k/ may also vary by speaker or dialect group, and still requires inferential reasoning about e.g. the position of the velum.

Another tack would be to interpolate a hypothesized central 'schwa' point (e.g. Proctor 2009), based on the overall dimensions of each speaker's articulatory space (as inferred from the full set of tongue surface tracings). This method gives a speaker-dependent, fixed reference point which is reliably 'central', allowing tongue fronting and backing to be understood as x-axis distance from the vertical centerline defined by 'schwa'. However, being based on tongue surface tracings, it still constitutes a rough approximation of a true anatomical landmark (the neutral, rest position of the tongue), and would thus not be exempt from worries of circularity. In addition, the calculation of the 'schwa' point is skewed by the fact that the tongue tip/blade may be poorly imaged. Lastly, it remains unclear how to test for statistically significant palatalization or velarization (i.e. deviation from the vertical centerline).

A third alternative is to divide the articulatory space into several articulatory regions with a grid. If the apex of the tongue body falls in a particular region, that segment can be considered to be palatalized (or velarized, central, etc.). Compared to point landmarks, this is a coarser metric. On the plus side, such grids avoid making claims about anatomical features, and give hard thresholds for classifying consonants as velarized/central/palatalized. However, the approach still needs to infer a centerline to construct the grid, and we lack a principled criterion for determining which regions should count as statistically significant palatalization/velarization.

A second issue we face is the question of what the displacement metric should track: a single point on the tongue body (the apex), or some larger portion of the tongue surface? Since we are interested in place of articulation contrasts, it makes sense to track a single point at the location of maximum constriction (the y-axis maximum of each tongue surface). But sometimes the tongue 'apex' (y-axis maximum) is at the tip/blade rather than the body of the tongue. Given that we're interested in fronting/backing of the *tongue body*, this introduces a major confound into our statistical method. This confound argues in favor of a more holistic, surface-based metric over a single-point measure. At present, we are comparing the relative merits of different surface-based metrics of this sort.

Kaytetye consonant contrasts without contours

Susan Lin, Benjamin Davies, and Katherine Demuth

Macquarie University

Two of the benefits of articulatory research using ultrasound technology are (a) the relative immediacy of access to visual information and (b) the portability of many ultrasound devices. Both points make ultrasound imaging an attractive option for field researchers interested in documenting lingual movement in remote locations.

Typical procedures for extracting and analysing data from ultrasound video involve at least 3 steps: extraction of individual frames from ultrasound video, followed by edge-detection of lingual contours, and calculation of difference and/or distance measures based on above contours. These steps, especially edge-detection, are often time-consuming and are often not suitable for quick visual comparisons while in the field. We describe an alternate approach (which we call "sighted" for the time being) potentially more suitable for such applications, and compare results using sighted values approach against results using more traditional measured values.

Procedures

The ultrasound images used in this comparison were extracted from ultrasound video recordings of native speakers of Kaytetye, an endangered Arandic language spoken near Alice Springs, NT, Australia. Like most Arandic languages, Kaytetye contains a four-way coronal place contrast – dental, alveolar, palatal, and retroflex – in oral stops, nasal stops, pre-stopped nasals, and laterals. For the purpose of this comparison, we targeted the articulatory differences between the dental stop /t̪/ and the alveolar stop /t/. Ultrasound video was collected from seven female speakers of Kaytetye, who produced 16-24 instances of the target consonants in either /a#_a/ or /a#_əə/ contexts. For each target sequence, two frames from the ultrasound video were manually selected:

1   Vocalic frame – one frame from production of the preceding vowel /a/, during lingual stability

2   Consonantal frame – one frame from production of the target consonant, at peak constriction

Sighted

In the sighted approach, the consonantal frame was tinted red, with all pixels below a specific black threshold made transparent, then superimposed on the vocalic frame. This process was repeated for each item, resulting in a single overlay image (as in Figure 1) for each target utterance. These images were given to a naive coder, who has seen lingual ultrasound images previously, but who has had no training in ultrasound speech research. The coder was instructed to "mentally trisect the white line in the centre of the image, and give each section a rating of +1/0/-1 if the red line is farther from/the same distance as/closer to the centre of the image."
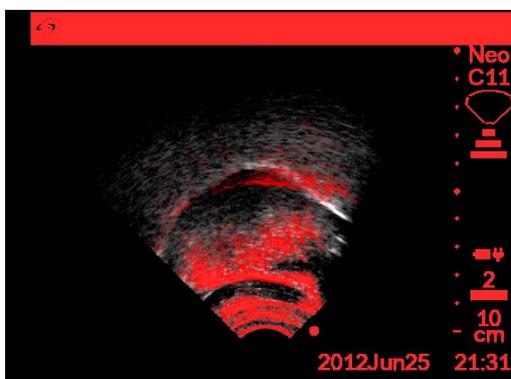

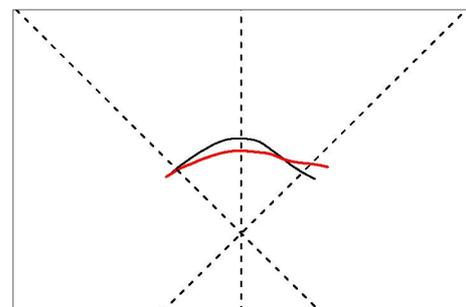**Figure 2. Vocalic /a/ (black) and consonantal /t̪/ (red) contours, with "back", "middle", and "front" intersect lines**

Figure 1. Vocalic /a/ (white) and consonantal /t̪ / (red) overlay.

Measured

For the more traditional measures, the lingual contour was traced in both the vocalic and consonantal frames, using EdgeTrak (Li et al. 2005). Three lines were created to approximate the trisection used in the sighted task (see Figure 2), and the distance (in pixels) between the black vocalic contour and the red consonantal contour were calculated. If the red consonantal contour was further from the point of origin of the three lines, the distance value was positive; otherwise, if it was closer to the point of origin, the distance value was negative, mirroring the polarity of the Height values from the sighted task.

Results

For the purpose of testing comparability of results, we asked the following question: does production of alveolar compared to dental stops in Kaytetye involve different motion of the tongue in the back/middle/front?

Measured results

A set of linear mixed effects models was run using the lme4 package in R (Bates et al. 2012) on the V-to-C Distance values, separately in each region (back, middle, front), with fixed factor Place (alveolar or dental), and random factors Subject, Word, and Repetition.

Figure 3 shows, for each region, the distribution of measured Distance between alveolar and dental stops. In the Back region, Distance between vocalic and consonantal contours were not significantly different between alveolar and dental places of articulation ($\beta = -0.30$, $t = -0.25$, $p = 0.8042$). In the Middle region, consonantal contours in dental stops were significantly lower compared to alveolar stops ($\beta = -13.34$, $t = -10.95$, $p < 0.0001$). In the Front, consonantal contours were again significantly lower in dental stops compared to alveolar stops, but the magnitude of difference was substantially smaller than in the Middle region ($\beta = -3.40$, $t = -2.32$, $p = 0.0217$).
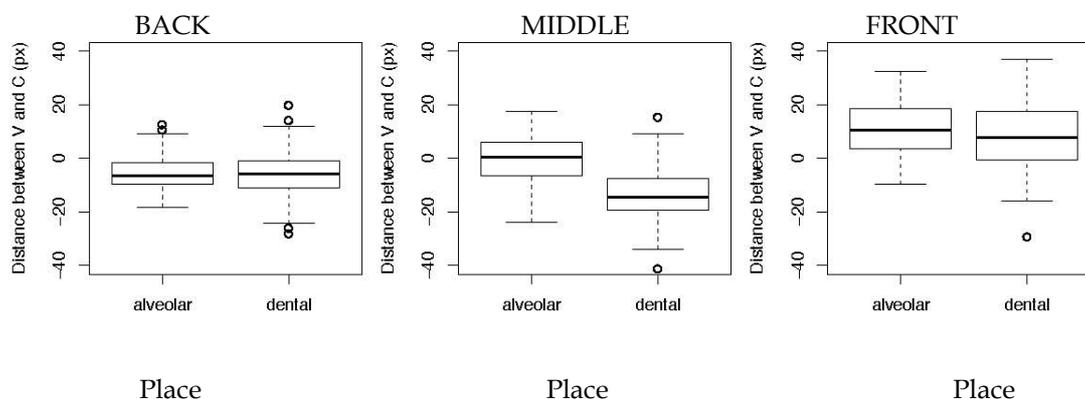

Figure 3. Distribution of measured V-to-C Distance values (px) by region and place of articulation.

Sighted results

Another set of linear mixed effects models was run the sighted Height values, again separately in each region, with fixed factor Place, and random factors Subject, Word, and Repetition.

Figure 4 shows, for each region, the proportion of sighted Height ratings between alveolar and dental stops. Results in the Back and Middle regions match the results for the measured values: Place did not have a significant effect on Height scores ($\beta = 4.08$, $z = -6.823$, $p = 0.4544$) in the Back region, but Height was significantly more likely to be negative in dental stops than alveolar stops ($\beta = -4.94$, $z = 6.82$, $p < 0.0001$). In the Front, Height scores were not significantly affected by Place ($\beta = -0.19$, $t = 0.002$, $p = 0.9980$), in contrast with the results from the measured

Measured v. sighted results

Table 1 displays the mean values for both Distance and Height, for this particular question. The sighted ratings concur with the measured distances for two of the three regions: Back and Middle. However, the difference in magnitude of tongue raising in the Front region was not picked up by the sighted

Height ratings. This is likely due to the sighted rating system outlined above not allowing for rating of different magnitudes, resulting in the vast majority of items, whether alveolar or dental, having a +1 (raising) rating in the Front region.
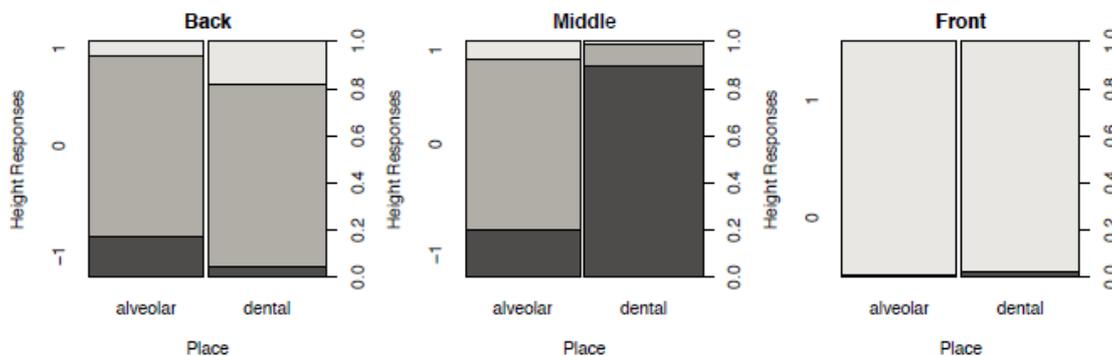

Figure 4. Proportion of -1 (dark gray), 0 (gray), and +1 (light gray) sighted Height ratings, by region and place of articulation.

	Sighted Means			Measured Means (px)		
	Back	Middle	Front	Back	Middle	Front
Alveolar	-0.11	-0.12	0.99	-5.02	-0.25	11.04
Dental	0.13	**-0.88**	0.98	-5.63	**-12.79**	8.35

Table 1. Mean scores across all subjects, for sighted Height (left) and measured Distance (right), by region and stop type. Significant differences (at p<0.05) between alveolar and dental highlighted in bold.

We also compared, more directly, the values from the two methods. Figure 5 shows, for each region, the measured Distances by Height rating. In this figure, values were aggregated across Place. We ran a linear mixed effects model, with measured Distance as the (continuous) dependent variable, and sighted Height as a (discrete 3-level) fixed factor. Subject, Iteration and Place were included as random factors. Again, the comparison was run independently for each region. For the discrete 3-level factor Height, 0 was set as the base value for comparison.
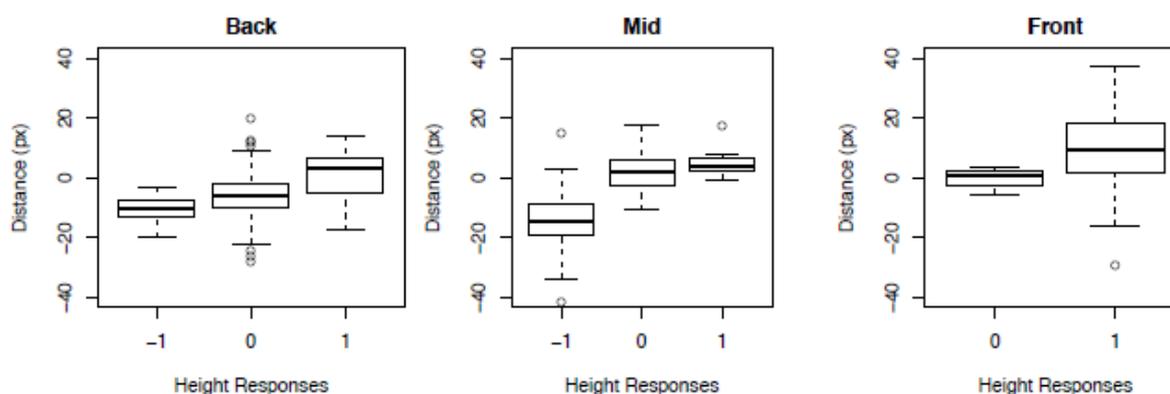
Figure 5. Measured Distance v. sighted Height responses, by region and place of articulation.

For the Back region, measured Distance was significantly lower when sighted Height was -1, compared to 0 (β = -4.16, t = -2.12, p = 0.0360), and significantly higher when Height was +1 (β = 5.88, t = 3.07, p = 0.0025). Similarly, for the Mid region, measured Distance was significantly lower when Height was -1 (β = -11.15, t = -7.11, p < 0.0001), and significantly higher when Height was +1 (β = 6.66, t = 2.49, p = 0.0138).

In contrast, for the Front region, sighted Height was not a significant predictor for measured Distance (β = 1.67, t = 0.31, p = 0.7583). It is quite plausible, however, that the lack of significant correlation between the two metrics in the Front region is due once again to the fact that exceedingly few items resulted in a 0 or -1 rating for Height – 3 of 165 total items, or 1.8% of all items.

Discussion

The sighted ratings approach appears to be most informative when differences are sufficiently large. There are several ways in which the very basic approach outlined here could be amended to result in measures that are more comparable to a more traditional measured approach.

In the sighted procedure outlined here, the coder was given extremely broad instruction as far as where to focus his attention. Some of the differences between polarity (+/-) between human coding and distance measures in our comparison are likely due to focus at different locations. While the measured procedures fixed location of measure, the sighted procedure allowed the coder to choose any region that he felt was sufficiently "forward," "backward," or "central." Restricting the area of focus visually for the human coder could be achieved using a reasonably simple procedure, such as adding opaque overlay to the combined image, to obscure visual access to a specific area.

Of course, one of the clear advantages of traditional measures, compared to the sighted procedure outlined in this abstract, is the availability and interpretability of differences in magnitude. To partially address this, the sighted procedure outlined here may be expanded to include more granularity. For instance, the coder may have five degrees of difference to choose from instead of three. Even so, the use of sighted procedures, as a whole, are likely not appropriate for issues that investigate very small differences.

References

Li, Min, Yusuf Akgul, & Chandra Kambhamettu. 2005. EdgeTrak [Computer Program]. Version 1.0.0.4. Retrieved October 3, 2008, from vims.cis.udel.edu/EdgeTrak/.

Bates, Douglas, Martin Maechler, & Ben Bolker. 2012. lme4: Linear mixed-effects models using S4 classes [R Package]. Version 0.999999-0. Retrieved July 10, 2012, from cran.r-project.org.

Session Three

The role of gesture timing in postvocalic /r/ lenition: an ultrasound study

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

[1]CASL, Queen Margaret University   [2]University of Glasgow

Researchers using articulatory analysis techniques have noted a lack of synchrony between the posterior and anterior gestures of coda liquids in English, e.g. (see Sproat and Fujimura, 1993, and Gick, Campbell, Oh and Tamburri-Watt, 2006). In these studies, posterior lingual gestures tended to occur closer to syllable centres, while anterior lingual gestures occurred closer to syllable margins. This asynchrony provides an articulatory mechanism that can explain vocalization, lenition and loss of coda liquids. We provide empirical evidence of how changes in gesture timing are contributing to coda /r/ lenition in a rhotic variety of English.

Since the late 1970s, sociolinguists have identified and studied postvocalic /r/ lenition in the Scottish Central Belt - the central lowland area of Scotland (see Romaine, 1978; Macafee, 1983; Stuart-Smith 2003, 2007; Lawson, Stuart-Smith and Scobbie, 2008, Jauriberry, Sock, Hamm and Pukli, 2012). Researchers carrying out auditory-acoustic analyses of postvocalic /r/ report difficulty in identifying a rhotic segment in words traditionally containing coda /r/ and describe a pharyngealised or velarized quality in prerhotic vowels, (e.g. see Speitel and Johnston, 1983; Stuart-Smith, 2007). They also show that these weakly rhotic or "derhoticised" postvocalic /r/ variants occur mainly in working-class speech, while

middle-class speakers use strongly rhotic postvocalic /r/ variants and exhibit different prerhotic vowel modifications.

In 2012, we collected a socially-stratified audio-ultrasound corpus of adolescent speech (16 speakers aged 12-13) in the city of Glasgow with equal numbers of working-class and middle-class participants and equal numbers of males and females. We collected word-list recordings from each speaker, containing 31 words with postvocalic /r/ after the full range of Scottish monophthongs, mixed in with a set of distractors. It was rarely possible to obtain an ultrasound image of the tongue that reliably showed the full extent of tongue root retraction, as well as the anterior /r/ gesture, so rather than looking at the temporal difference between the posterior and anterior /r/ lingual gestures, we focused on the temporal difference between the point of maximum constriction of the anterior lingual gesture (annotated as "rmax") and the offset of voicing in the syllable. We carried out a comparative analysis of anterior gesture timing in postvocalic /r/ word tokens obtained from working-class and middle-class speakers in the corpus. /r/ word tokens were also rated by the authors on a 7-point auditory rhoticity index using a Praat multiple-forced-choice interface that presented tokens in random order. We aimed to find out if *rmax* occurred later in the syllable in working-class speech than in middle-class speech, and, in particular, if working-class *rmax* tended to occur *after* the offset of voicing. We also wanted to find out if a delayed *rmax* gesture correlated with a weak percept of rhoticity.

Our analysis showed a significant difference between the timing of *rmax* in the working-class and middle-class tokens of post-vocalic /r/, with *rmax* occurring, on average, before the offset of voicing in middle-class dataset and after the offset of voicing in working-class dataset. There was also a significant negative correlation between *rmax* timing and rhoticity index score, i.e. delayed *rmax* timing in the syllable corresponded to a low rating on the rhoticity index, while an early *rmax* timing corresponded to a high rhoticity rating.

Our results provide empirical evidence that the lenition of postvocalic /r/ here does not result from a weakening of the gestures involved in /r/ production, i.e. gestural undershoot, but a change in the timing of the gestures involved in /r/ production in relation to voicing offset. Even though many audio tokens in the study were rated as "/r/-less" on the rhoticity index, the articulatory data showed that /r/ gestures were present, but delayed. The presence of a delayed anterior /r/ gesture helps to explain the difficulties encountered in previous auditory-acoustic studies in classifying weakly rhotic variants encountered in Central Belt speech, as it is probable that some, but not all, of the /r/ articulation occurred before the offset of voicing, leaving the /r/ partly or wholly inaudible.

References

Gick, Bryan, Campbell, Fiona, Oh, Sunyoung. and Tamburri-Watt, Linda (2006) 'Toward universals in the gestural organization of syllables: a cross-linguistic study of liquids', *Journal of Phonetics*, 34, 49-72.

Jauriberry, T. Sock, R. Hamm, A. and Pukli, M. (2012) "Rhoticité et dérhoticisation en anglais écossais d'Ayrshire". *Actes de la conférence conjointe JEP-TALN-RECITAL*, Grenoble 4-8 June 2012 (vol. 1) pp89-96

Lawson, E. Stuart-Smith, J. & Scobbie, J. M. (2008) "Articulatory insights into language variation and change: Preliminary findings from an ultrasound study of derhoticisation." In *Selected Papers from NWAV 36, Special Issue of Pennsylvania Working Papers in Linguistics* (36). Pennsylvania: The Penn Linguistics Club, pp102-110.

Macafee, C. (1983) *Glasgow. Varieties of English Around the World. Text Series*, Vol.3. Amsterdam: John Benjamins Publishing Co.

Romaine, S. (1979) "Postvocalic /r/ in Scottish English: Sound change in progress?" In P. Trudgill, (Ed). *Sociolinguistic Patterns in British English.* London: Edward Arnold, pp144-157.

Speitel, Hans H. and Johnston, Paul, A. (1983) ESRC End of Grant Report 'A Sociolinguistic Investigation of Edinburgh Speech.'

Sproat, Richard and Fujimura, Osamu (1993) Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, 21, 291-311.

Stuart-Smith, J. (2007) "A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents." *Proceedings of the XVIth International Congress of Phonetic Sciences*, Saarbrücken, 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.* Edinburgh: Edinburgh University Press, pp 110-137.

Ultrasound+corpus study of phonetically-motivated variation in North American French and English

Jeff Mielke and Christopher Carignan

North Carolina State University

This presentation reports two projects employing ultrasound imaging to augment conversational speech corpora of French from Gatineau, Quebec, and English from Raleigh, North Carolina. Targeted articulatory data are collected in order to study phonetically-motivated sound changes in progress. While phonetically-motivated change from below is a fundamental concept in contemporary approaches to phonology and variation, empirical data is sparse (Cedergren, 1973; Trudgill, 1974; Labov, 2001), partly because changes usually go unnoticed until long after their inception, and because articulatory data on changes in progress (which can shed light on phonetic motivations) is often unavailable (cf. Scobbie et al. 2008).

The Gatineau project investigates the development of rhotic vowels in Canadian French. Some young speakers of Canadian French in Southwestern Quebec and Eastern Ontario produce the vowels /ø/, /œ/, and /œ̃/, making them sound like English [ɹ] (i.e., *heureux*, *docteur*, and *commu*n sound like [ɹ R ɹ̩], [dɔktʌʀ], and [kɔmɹ̩̃]). When asked, native speakers are completely unaware of the difference between rhotic and non-rhotic pronunciations, suggesting that rhoticity is a change from below. Previous reports of retroflex-sounding variants of Canadian French vowels date back to the early 1970s in Montreal (Dumas 1972, 100, Sankoff, p.c.), and a retroflex-sounding variant of /ʀ/ has also been observed (Sankoff and Blondeau, 2007). This change is of special interest because it resembles North American English /ɹ/, which can be produced with various tongue shapes, including bunched and retroflex variants (Delattre and Freeman, 1968), raising the question of whether French rhotic vowels are also produced with these categorically different tongue shapes.

The historical development and distribution of rhotic vowels were investigated by acoustic analysis of data from 75 speakers born between 1893 and 1991, from two corpora of Ottawa-Hull French: *Corpus du français parlé à Ottawa-Hull* (Poplack, 1989) and *Corpus du fran̡çais de l'Outaouais au nouveau millénaire* (Poplack and Bourdages, 2010). Reduction in F3 values (i.e., rhoticity) for word-final /ø/ and /œ̃/ was observed for speakers born after 1960 (Figure 1, left). Rhotic /œ/ before /ʀ/ appears somewhat later. Importantly, F3 lowering (~rhoticity) emerged gradually, beginning in speakers born after 1965, and rhoticity was preceded by F2 lowering (~tongue retraction or lip rounding).

Mid-sagittal ultrasound data from 23 native speakers of Canadian French revealed that the most rhotic speakers use bunched and retroflex tongue shapes at rates comparable to those seen for English /ɹ/ in similar contexts. Moderately rhotic vowels were produced with tongue shapes ranging from bunched to vowel-like. Figure 1 (right) shows examples of tongue shapes for two extremely rhotic speakers. One young speaker who produced extremely rhotic vowels produced them with retroflex tongue shapes, suggesting that the retroflex tongue posture emerged only after the change progressed to the point where the perceptual target was extreme enough to motivate a learner to use a non-vowel-like tongue posture.

The second project draws participants directly from a 200-speaker corpus of Raleigh, NC English (Dodsworth and Kohn, 2012). Data collection for this project is still in progress. The variables under investigation are /æ/ tensing, postlexical /s/-retraction, /ɹ/ flapping, intrusive [l], /l/ vocalization, and pre-liquid vowel changes. While these variables differ in their salience, they are all expected to be sensitive to inter-speaker articulatory variation in tongue and velum movements.
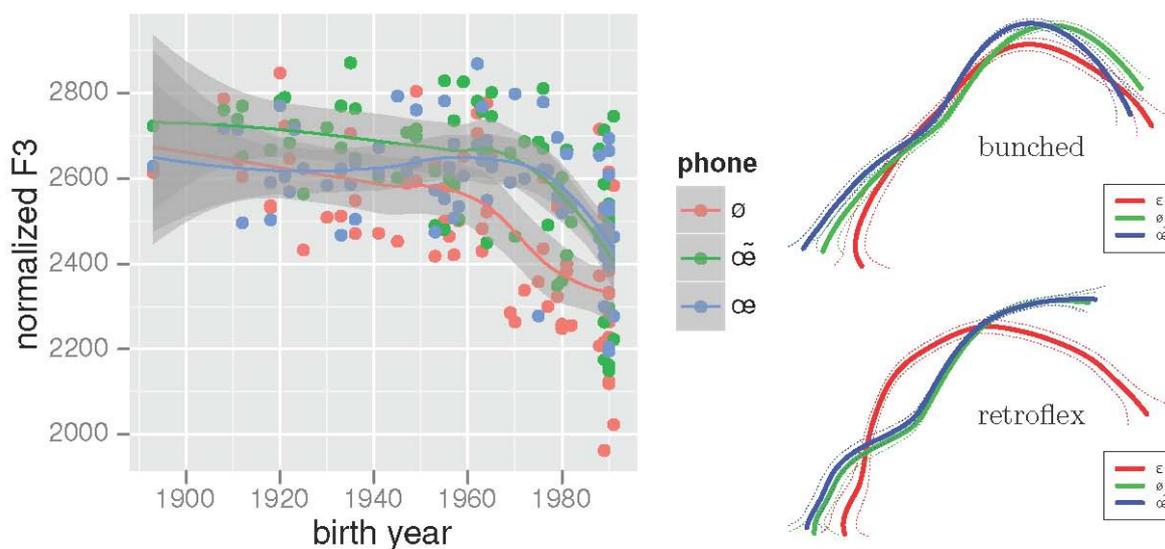
Figure 1: Left: Gatineau French mid front rounded vowel F3 by birth year (speaker means);
Right: Tongue contours for two speakers' rhotic vowels, with non-rhotic vowel [ɛ] for reference

References

Cedergren, Henrietta. 1973. The interplay of social and linguistic factors in Panama. Doctoral Dissertation, Cornell University.

Delattre, Pierre, and Donald C. Freeman. 1968. A dialect study of American r's by x-ray motion picture. *Linguistics* 44:29–68.

Dodsworth, Robin, and Mary Kohn. 2012. Urban rejection of the vernacular: The SVS undone. *Language Variation and Change* 24:221–245.

Dumas, Denis. 1972. Le français populaire de Montréal: description phonologique. Universit´e de Montréal M.A. thesis.

Labov, William. 2001. *Principles of linguistic change: Social factors*. Oxford: Blackwell.

Poplack, Shana. 1989. The care and handling of a mega-corpus. In *Language change and variation*, ed. Ralph Fasold and Deborah Schiffrin. Amsterdam: John Benjamins.

Poplack, Shana, and Johanne S. Bourdages. 2010. Normes et variation en français : Rivitalité entre l'école, la communauté et l'idéologie. Research project funded by SSHRC (# 410-2005-2108 [2005-2008]), Ottawa, Ontario.

Sankoff, Gillian, and Hélène Blondeau. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83:560–588.

Scobbie, James, Jane Stuart-Smith, and Eleanor Lawson. 2008. Looking variation and change in the mouth: Developing the sociolinguistic potential of ultrasound tongue imaging. Unpublished Economic and Social Research Council (ESRC) end of grant report, Swindon, U.K.

Trudgill, Peter, ed. 1974. *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.

An Ultrasound and Acoustic Study of Consonant Cluster Simplification in Persian

Reza Falahati

University of Ottawa

The main goal of this study is to investigate the nature of an optional consonant deletion process, through an articulatory and acoustic study of word-final consonant clusters in Persian. Persian word-

final coronal stops are optionally deleted when they are preceded by obstruents or the homorganic nasal /n/. For example, the final clusters in the words /næft/ "oil", /suχt/ "burnt" and /qæsd/ "intention" are optionally simplified in fast/casual speech, resulting in: [næf], [suχ], and [qæs]. What is not clear from this traditional description is whether the coronal stop is truly deleted, or if a coronal gesture is produced, but not heard, because it is obscured by the adjacent consonants. According to Articulatory Phonology (Browman & Goldstein 1986, 1988, 1989, 1990a, 1990b, 1992, 2001), the articulatory gestures of the deleted segments can still exist even if the segments are not heard. In this study, ultrasound imaging was used to determine whether coronal consonant deletion in Persian is categorical or gradient, and the acoustic consequences of cluster simplification were investigated through duration and spectral measures. This phonetic study enables an account for the optional nature of the cluster simplification process.

Ten Persian-speaking graduate students from the University of Ottawa and Carleton University, five male and five female, aged 25-38 participated in the articulatory and acoustic study. Audio and real time ultrasound video recordings were made while subjects had a guided conversation with a native speaker of Persian. 662 tokens of word-final coronal clusters were auditorily classified into unsimplified and simplified according to whether they contained an audible [t]. Singleton coda consonants and singleton /t/s were also captured as controls. The end of the constriction plateau of C1 and beginning of constriction plateau of C3 were used to define a time interval in which to measure the coronal gesture as the vertical distance between the tongue blade and the palate. Smoothing Splines ANOVA was used in a novel way to compare tongue blade height over time across the three conditions.

The articulatory results of this study showed that the gestures of the deleted segments are often still present. More specifically, the findings showed that of the clusters that sounded simplified, some truly had no [t] gesture, some had gestural overlap, and some had reduced gestures. In order to explain the optional nature of the simplification process, it is argued that the simplified tokens are the result of two independent mechanisms. Inevitable mechanical and physiological effects generate gesturally reduced and overlapped tokens whereas planned language-specific behaviors driven by phonological rules or abstract cognitive representations result in no [t]-gesture output. The findings of this study support the main arguments presented in Articulatory Phonology regarding the underlying reasons for sound patterns and sound change. The results of this study are further used to examine different sound change models. It is argued that the simplified tokens with totally deleted [t] gesture could be the result of speakers changing their representations based on other people's gestural overlap. This would be instances of the Choice and Chance categories in Blevins' (2004) CCC sound change model. The acoustic results did not find any major cues which could distinguish simplified tokens from controls. It is argued that articulatory data should form an integral part of phonetic studies.

Session Four

Laryngeal contributions to weak vs. strong ejectives

Sonya Bird, Scott Moisik, John Esling, Peter Jacobs

University of Victoria

In their description of ejective stops, Ladefoged & Johnson (2011) note that a crucial component of their articulation is *raising* of the closed larynx: this gesture increases the pressure build-up behind the oral closure, ensuring the characteristic high-amplitude burst as the oral closure is released. While such 'strong' (or 'tense') ejectives have been attested in some languages (Lindau 1984 and others), so have 'weak' (or 'lax') ejectives, which do not seem to fit this description (Kingston 1985; Warner 1996; Davis & Hargus 1999). A number of acoustic studies have been conducted on ejectives across languages of the world (Kingston 1985; Lindau 1984; Warner 1996; Davis & Hargus 1999; Bird 2002). Based on these, a fair amount is known about the acoustic differences between strong and weak ejectives, in terms of VOT, amplitude rise time into the following vowel, F0 and voice quality at onset of the following vowel. Much less is known about the articulatory patterns which underlie these acoustic

differences. Kingston (1985) hypotheses that the strong-weak distinction is based on vocal fold stiffness rather than larynx height. However, this view does not consider extraglottal mechanisms available for strong-weak differentiation. For example, palatopharyngeus action could drive pharyngeal lumen reduction (e.g. Zemlin 1998) in strong ejectives, and a moderate engagement of the epilaryngeal mechanism (Moisik 2013) could support the production of weak ejectives. More work is clearly needed to understand the articulatory properties of ejectives.

This paper reports on a pilot study designed to investigate what role larynx raising plays, if any, in creating the difference between strong and weak ejectives. Laryngeal ultrasound data were collected from three trained phoneticians and a single fluent speaker of Skwxwu7mesh (Squamish; Central Salish)[1]. The trained phoneticians produced strong and weak ejectives, as well as voiceless aspirated stops (for comparison), at labial, alveolar, velar, and uvular places of articulation, and in both #_V and V_V contexts (where V = [i a u]). The strong ejectives were produced by imitating those of Lakhota (Ladefoged & Johnson 2011: audio files); the weak ejectives by imitating those of Dakelh (Bird, field recordings). Imitation was used to provide an auditory model to aim for, and hence to minimize production bias. The Skwxwu7mesh speaker produced ejectives embedded in real words, at all possible places of articulation - [p' t' k' kʷ' q' qʷ'] - and a variety of vowel contexts. The following table summarizes the laryngeal contribution to different types of stops, based on visual inspection of the ultrasound videos as well as optical flow analyses (Moisik et al. 2011).

Stop Type	Direction of larynx displacement	Amplitude of larynx displacement	Velocity of larynx displacement
Voiceless aspirated	down or up	Slight	Slow
Strong ejectives	(down-)up-down	**Great**	**Fast**
Weak ejectives	up	Slight	Slow

Larynx displacement in strong ejectives - including those of Skwxwu7mesh - were particularly striking, in terms of the properties listed in the table above but also in terms of *timing*: the larynx seemed to lower during the stop, and spring back up very quickly just before the release, reminiscent of the trajectory of a slingshot. These results indicate that, contrary to Kingston's (1985) claim, larynx height *is* involved in creating the distinction between strong and weak ejectives. They also show that the timing of the laryngeal gesture(s) with respect to the oral one(s) is perhaps not as tightly bound as Kingston (1985) suggested, but rather can be manipulated by speakers of a language in realizing the sounds in their inventories.

References

Bird, S. (2002). Dakelh ejectives: Evidence for new ways of classifying sounds. Paper presented at the 76th Linguistic Society of America, San Francisco CA.

Davis, C. & S. Hargus (1999). *The glottal contrast in Witsuwit'en*. Unpublished manuscript. University of Washington.

Kingston, J.C. (1985). *The phonetics and phonology of the timing of oral and glottal events*. Dissertation. University of California.

Ladefoged, P. and K. Johnson (2011). *A course in Phonetics, Sixth Edition*. Boston MA: Wadsworth, Cengage Learning.

Lindau, M. (1984). Phonetic differences in glottalic consonants. *Journal of Phonetics* **12**:147-155.

Moisik, S. R. (2013). *The epilarynx in speech*. Dissertation. University of Victoria, BC.

Moisik, S. R., Esling, J. H., Bird, S., & Lin, H. (2011). Evaluating laryngeal ultrasound to study larynx state and height. In W. S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 136–139). Hong Kong.

Warner, N. (1996). *Acoustic characteristics of ejectives in Ingush*. Proceedings of the International Conference on Spoken Language Processing, October 3-6, 1996. Philadelphia, PA. 1525-1528.

Zemlin, W.R. (1998). *Speech and hearing science: Anatomy and physiology, Fourth Edition*. Boston MA: Allyn and Bacon.

---

[1] All speakers were co-investigators on the project.

Morphological effects on the darkness of English intervocalic /l/

Sang-Im Lee Kim and Lisa Davidson

New York University

Articulatory and acoustic studies have provided evidence that in word-initial and word-final positions, English /l/ exhibits substantial differences in 'darkness': darker [ɫ] in word-final position is produced with a more retracted tongue dorsum and lowered tongue body than lighter [l] in word-initial position. The darkness of intervocalic /l/, however, is variable. Using articulatory and acoustic evidence, Sproat and Fujimura (1993) argue that /l/ darkness is on a continuum affected by duration and boundary strength. Hayes (2000) claims that the morphological status of intervocalic /l/s should affect whether they are produced as light or dark variants. He distinguishes between two types of intervocalic /l/s which would be syllabified as an onset by most phonological theories: those where the /l/ is the final consonant of a related monomorpheme and found before a morpheme boundary (*feel#ing*), and those where /l/ is part of the suffix after the boundary (*free#ly*). Unlike Sproat and Fujimura, Hayes argues that there are discrete light and dark allophones of /l/, and that dark [ɫ] should be produced more often when it is before the morpheme boundary and that light [l] should be found more frequently after the boundary. In this study, ultrasound imaging is used to investigate whether the morphological affiliation of the /l/ affects the darkness of /l/.

Departing from Hayes (2000), we predict that complex morphological and phonological factors give rise to gradient darkness for the following three types of stimuli: (1) post-boundary: when /l/ corresponds with the onset of a suffix (e.g. *flaw#less, coup#less*), (2) pre-boundary: when /l/ corresponds with the final position of the stem word (e.g. *tall#est, cool#est*), and (3) stem: when /l/ is the final consonant of a stem word (e.g. *tall, cool*). In particular, pre-boundary /l/ is predicted to be darker than the post-boundary /l/ due to the potential paradigmatic relation with the coda /l/ in the stem form, whereas it is predicted to be lighter than the stem /l/ due to possible phonological pressure for there to be an onset for the following vowel-initial suffix. In the ultrasound imaging, six American English speakers produced these three types of stimuli in various vowel contexts. The words were embedded in short phrases (e.g. 'the tallest building', 'the coupless revolution'), randomized, and repeated 10 times.

To compare the degree of retraction among the three different types of /l/, the frame corresponding to the most retracted position of the tongue dorsum during the production of the /l/ was selected. The articulation of /l/ was compared using a smoothing spline ANOVA (SS ANOVA, Davidson 2006, Wahba 1990). In Figure 1, comparing the most retracted frame of the /l/s in *coupless* (in grey) and *coolest* (in dark), the top figure represent each repetition for one speaker and the bottom are the splines which indicate the best fit to the data. The tongue curves were divided into thirds where the leftmost portion roughly corresponds to the tongue dorsum (TD) and the middle portion corresponds to the tongue body (TB). In this particular case, the pre-boundary /l/ (*coolest*) was coded as more retracted for TD and lowered for TB than the post-boundary /l/ (*coupless*). For the acoustic analysis, F1 and F2 were measured where the maximal retraction frame was captured, and normalized intensity was also calcuated.

The results showed that the predicted order was upheld for both articulatory and acoustic measures. For most cases, either tongue dorsum retraction or tongue body lowering indicated greatest darkness for the stem /l/ and intermediate for the pre-boundary /l/, followed by the post-boundary condition. Likewise, F1, normalized intensity, and to a lesser extent F2 measures were also consistent with the articulatory results. The results are consistent with Hayes (2000) in that speakers produce a darker /l/ for the pre-boundary /l/ to maintain the paradigmatic relationship with the stem /l/. However, the results crucially refute the assumption of a strict bimodal allophonic production for /l/. Rather, the presence of the intermediate implementation for the pre-boundary /l/ is compatible with /l/ in the phonetic continuum proposed by Sproat and Fujimura (1993). We also discuss a challenge that our data presents for most current formal phonological theories where generating gradient forms is not trivial.
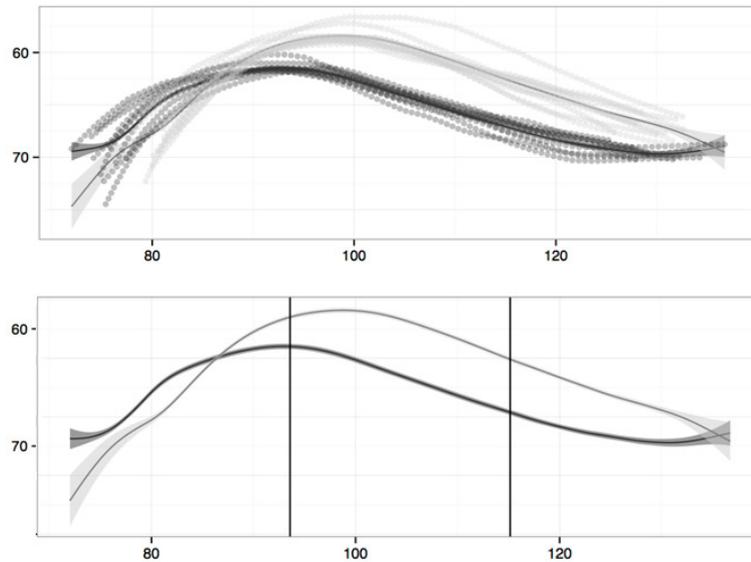
Figure 1. Top. Raw data points from ten repetitions (dotted lines) for comparison of the tongue shape for /l/ in *coolest* (in black) vs. *coupless* (in grey) for subject AD**. Bottom.** Smoothing spline estimate and 95% Bayesian confidence interval for comparison of the mean curves. The vertical lines are a rough division of the vocal tract into tongue dorsum (left), tongue body (middle), and tongue tip/blade (front). The tongue tip is on the right and the tongue dorsum is on the left.
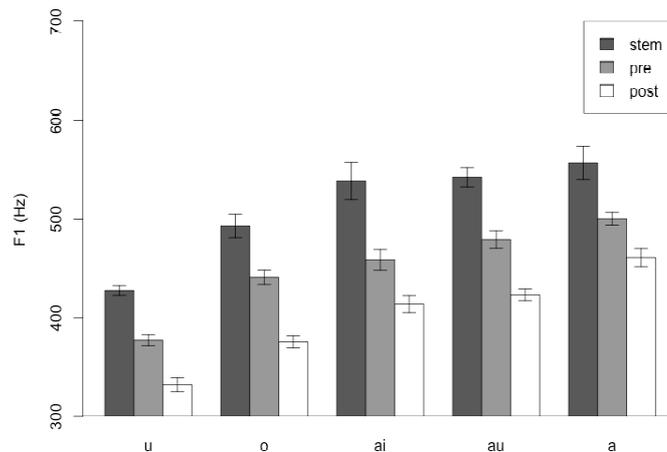

Figure 2. Mean of F1 (Hz) at minimum F2. Error bars indicate standard error.

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.

Hayes, Bruce. 2000. Gradient well-formedness in Optimality Theory. In Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer, Eds., *Optimality Theory: Phonology, Syntax, and Acquisition*. Oxford: Oxford University Press, pp. 88-120.

Sproat, R. and Osamu Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21, 291-311.

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

An ultrasound study of retroflex and alveolar laterals in Kannada

Alexei Kochetov, N. Sreedevi, Midula Kasim and R. Manjula

Department of Linguistics, University of Toronto

Retroflex consonants in Dravidian languages are known to be produced with the underside of the tongue making a constriction behind the alveolar ridge (Ladefoged & Bhaskararao, 1983). The retraction and raising of the tongue blade is accompanied by some hollowing of the anterior tongue body, resulting in the characteristic concave tongue shape. Less is known, however, about the shape and the relative position of the posterior tongue body/root during the retroflex production. In a recent ultrasound study of Malayalam liquids, Scobbie et al. (in press) found that the retroflex lateral /ɭ/ was produced with a substantial retraction of the tongue root (pharyngealization), relative to alveolar /l/ and the preceding vowel /a/. An MRI study of Tamil liquids by Narayanan et al. (1999), however, revealed a wider pharyngeal cavity for /ɭ/ than /l/, indicative of some tongue root/body advancement for the former. Similar evidence for the retroflex tongue body advancement was found in an ultrasound study of Kannada stops by Kochetov et al. (2012), where the posterior tongue body for /ʈ/ was more front relative to the rest position and /t/, yet similar to that for /ʧ/. The partly contradictory results of these studies could be due to language-specific differences in the production of retroflexes, or due to articulatory differences between stops and laterals.

To address the issue, this paper presents results of an ultrasound study of laterals /ɭ/ and /l/ produced by 10 native speakers of Kannada (5 females and 5 males) from Mysore, Karnataka (South India). The materials included the words /haɭɭa/ 'small stream', /alla/ 'not', and /aʈʈa/ 'garret' (as a control), among other items. Ten repetitions of each word were collected using a PI 7.5 MHz *SeeMore* probe (*Interson Corporation*) connected via USB to a laptop computer. The participants wore an *Articulate Instruments* stabilizing headset. Tongue contours during the consonant closure were traced using *EdgeTrak* (Li et al., 2005). To evaluate differences in the tongue body position and shape between pairs of consonants, a series of *Smoothing Spline ANOVAs* (*SS-ANOVA*; Davidson, 2006) was performed, separately for each speaker and across the speakers. Analysis of the simultaneously collected acoustic data involved measurements of formants F1-F4 during, before, and after the lateral closure.

The results revealed that, for all 10 speakers, the posterior tongue body (and possibly the root) for /ɭ/ was as advanced as or significantly more advanced than for /l/ (see Figure 1, left). This indicated that /ɭ/-retroflexion was accompanied by some fronting of the tongue body. Acoustically, this was manifested in the consistently higher F2 for /ɭ/ than for /l/ (by 150-300 Hz). A comparison of the tongue shapes for /ɭ/ and /ʈ/ showed no significant differences in the degree of the tongue body fronting (see Figure 1, right). This showed that the more advanced tongue body position is not unique to /ɭ/, but is likely part of the production of Kannada retroflexes in general.

Combined with other studies, the current results point to language-particular differences in the realization of Dravidian laterals: /ɭ/ in Kannada is similar to /ɭ/ in Tamil (the fronted tongue body/root), and different from /ɭ/ in Malayalam (the backed tongue body/root). These articulatory differences, in turn, are consistent with language-specific acoustic differences in laterals: higher F2 for /ɭ/ than /l/ in Kannada (our data) and Tamil (Narayanan et al., 1999), while lower F2 for /ɭ/ than /l/ in Malayalam (Punnoose, 2011).
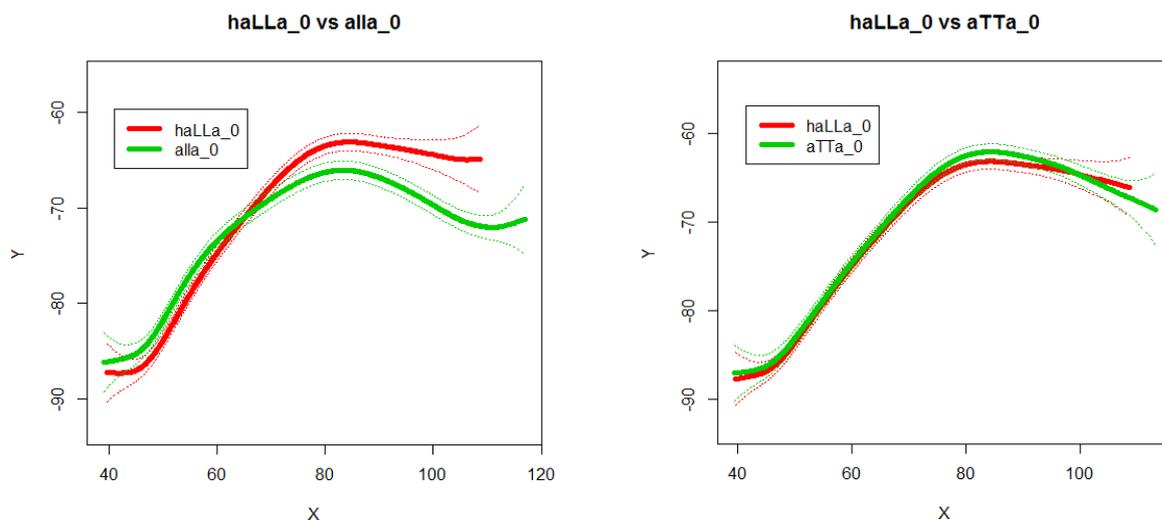
Figure 1. Results of SS-ANOVA comparisons of the tongue shapes for /l̪/ (haLLa) vs. /l/ (alla) and for /l̪/ (haLLa) and /ʈ/ (aTTa), performed across 10 speakers.

References

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.

Kochetov, A., Sreedevi, N., Kasim, M. & Manjula, R. (2012). A pilot ultrasound study of Kannada lingual articulations. *Journal of Indian Speech and Hearing Association*, *26,* 38–49.

Ladefoged, P., & Bhaskararao, P. (1983). Non-quantal aspects of consonant production: a study of retroflex sounds. *Journal of Phonetics*, *11*, 291–302.

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

Narayanan, S., Byrd, D. & Kaun, A. (1999). Geometry, kinematics, and acoustics of Tamil liquid consonants. *Journal of the Acoustical Society of America*, *106*, 1993–2007.

Punnoose, R. 2011. An auditory and acoustic study of liquids in Malayalam. University of Newcastle PhD thesis.

Scobbie, J., Punnoose, R., & Khattab, G. (in press). Articulating five liquids: a single speaker ultrasound study of Malayalam. In Rhotics: New data and perspectives. BU Press, Bozen-Bolzano.

Thursday 7th November	Session Five

Comparing principal component analysis of ultrasound images with contour analyses in a study of tongue body control during German coronals

Marianne Pouplier and Philip Hoole

Institute of Phonetics, LMU Munich

We have recorded data on German coronal consonants (/t, d, n, s, z, l/) in different vowel contexts in order to investigate to what extent coronals differ in tongue body position and vowel-context conditioned variability. In our recent modeling work on German coronals we found that all coronals need to be specified for a tongue body articulatory target in order to be able to synthesize appropriate CV and VC coarticulation patterns. This view is supported by German EMA data of Geumann et al

(1999) in which the non-sibilant coronals did not differ from each other in terms of vowel-context induced dorsal variability. Notably German clear /l/ patterned with the obstruents /t, d, n/, while the sibilant was more constrained. Mermelstein (1973) likewise argued for a tongue body target during coronals due to intervocalic coronal stops deflecting the V-to-V trajectory towards a consonantal target. This contrasts with Proctor's (2009) recent claim that /l/ but not the other coronals should cross-linguistically be specified for a tongue body gesture. He found for Spanish that tongue body during clear /l/ is less variable in different vowel contexts compared to coronal stops and concludes that tongue body control is cross-linguistically a feature differentiating liquids from other coronals.

In order to shed light on the articulatory synergies used for the production of coronals, we investigate tongue body position and vowel-context induced variability in our data for German. In doing so, we compare several analysis methods for ultrasound data. For one, we have traced the outlines of the tongue shape contour for each token and subject these contours to variability analyses based among others on the area between curves. Secondly, we explore the possibility of quantifying the change of pixel patterns over time in the ultrasound image directly based on principal component analysis (Hueber et al. 2007).

References:

Geumann, A., Kroos, C. and Tillmann, H.G. (1999) Are there compensatory effects in natural speech?, Proceedings of the 14th International Congress of the Phonetic Sciences, San Francisco, Vol. 1 399- 402.

Proctor, M. (2009). Gestural Characterization of a Phonological Class: The Liquids, PhD dissertation,Yale University.

Mermelstein, P. (1973). Articulatory model for the study of speech production, Journal of the Acoustical Society of America 53/ 4, 1070–1082.

Hueber, T., Aversano, G., Chollet, G., Denby, B., Dreyfus, G., Oussar, Y., Roussel, P., Stone, M., 2007. Eigentongue feature extraction for an ultrasound-based silent speech interface, IEEE International Conference on Acoustic, Speech, and Signal Processing, ICASSP07, Honolulu, vol. 1, pp. 1245-1248.

Lingual coarticulation and articulatory constraints in adolescents

Natalia Zharkova

Clinical Audiology, Speech and Language Research Centre, Queen Margaret University

This study describes lingual coarticulatory patterns in the consonants /p/, /t/, /s/, /ʃ/ in adolescents aged 13 years old. The Degree of Articulatory Constraint (DAC) model of lingual coarticulation (Recasens et al. 1997) employs constraints on the tongue as factors determining coarticulatory patterns displayed by speech segments, with more constrained tongue dorsum resulting in more resistance to lingual coarticulation, and higher DAC values. During bilabial consonants the tongue is almost completely unconstrained, and the whole tongue contour can adapt to the vocalic influence. During postalveolar consonants, the tongue predorsum is very constrained, hence they strongly resist to vowel-induced lingual coarticulation. In alveolar consonants, the tongue is less free to adapt to the vowel influence than in labials, but more than in postalveolars (Recasens 1999). Indeed, studies of adult speech have demonstrated more lingual coarticulation in /s/ than in /ʃ/ (Recasens & Espinosa 2009; Pouplier et al. 2011; Niebuhr & Meunier 2011; Zharkova et al. accepted). As opposed to the alveolar stop /t/, the fricative /s/ has an additional constraint on tongue sides, thus increasing its potential to resist the vocalic influence, compared with the stop (Stone et al. 1992).

Previous work has reported developmental shifts in the degree of lingual coarticulation of individual speech sounds. For example, for 10-12-year-old children, similarly to findings from adult speech, /s/ showed more coarticulation than /ʃ/ at mid-consonant, but this was not the case for 6-9-year-olds

(Zharkova et al. 2011; 2012; accepted). In order to understand the nature of the changes in consonant-specific coarticulatory patterns with age, it is necessary to trace coarticulation development throughout childhood in tightly defined age groups, and with the data including several consonants. Such a project is currently underway, and as part of it, the current study documented the amount of vowel-on-consonant coarticulation in CV syllables in 13-year-olds, using consonants which differ in their DAC properties. The amount of anticipatory vowel-on-consonant coarticulation for /p/, /t/, /s/ and /ʃ/ was predicted to decrease in this order, consistent with the DAC model.

The study used ultrasound tongue movement data recorded at the frame rate of 100 Hz and synchronised with the acoustic signal. CV syllables with the four consonants mentioned above and the vowels /a/ or /i/ were produced in the carrier phrase "It's a ..., Pam" (each target repeated five times) by ten speakers of Scottish Standard English, aged between 13 years 0 months and 13 years 11 months. Participants wore a headset for stabilising the ultrasound transducer in relation to the head (Articulate Instruments Ltd 2008). In each CV token, an annotation was placed on the display at mid-consonant (mid-closure for the stops), based on the acoustic data. Cubic splines were then fitted to midsagittal tongue curves at each annotation.

Presence and size of coarticulatory effects were established using the Nearest Neighbour method for comparing sets of tongue curves for a segment in contrasting segmental contexts (Zharkova & Hewlett 2009). The presence of a coarticulatory effect was determined by comparing the mean distance between two sets of tongue curves for the consonant (one set in the context of /a/ and another set in the context of /i/; 25 distance values in total) and the mean distances between the curves within each vowel context (10 distance values per vowel context). If the distance across the two sets of curves was significantly greater than both within-set distances then the coarticulatory effect on the consonant was deemed to be present. The size of effect was represented by the mean across-set distance. Linear mixed models with speaker as a random factor were performed in R, with *lmer* software package (Baayen 2008). Tukey post hoc tests were used to compare the coarticulatory effect size across consonants.

There was a significant effect from the vowels on all consonants. Mean sizes of coarticulatory effect across all speakers were as follows: 4.68 mm for /p/, 2.45 mm for /t/, 2.32 mm for /s/ and 1.43 mm for /ʃ/. A significant difference across consonants in size of coarticulatory effect was reported ($F = 173.62$). The size of the coarticulatory effect on /p/ was significantly greater than that on all the other consonants ($p < 0.001$). The posthoc tests showed a significant difference ($p < 0.001$) between /ʃ/ and /t/ and between /ʃ/ and /s/. The difference between /t/ and /s/ was not significant.

The consonants patterned as predicted in relation to the size of the vowel-induced effect, with the bilabial consonant being most affected by the vowel influence and the postalveolar being least affected. Similarly to previous findings on adult speech, /ʃ/ exhibited significantly less coarticulation than /s/. The alveolar fricative had less coarticulation than the alveolar stop, but the difference between the two was not significant. The latter result matched the pattern reported in a study of lingual coarticulation in adult speakers of Scottish Standard English (Zharkova 2008), where the extent of vowel-induced coarticulation was not much larger for /t/ than for /s/. In summary, these results show that the adolescents have developed an adult-like relation between the coarticulatory properties of the four consonants.

References

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

Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Niebuhr, O., & Meunier, C. (2011). The phonetic manifestation of French /s#ʃ/ and /ʃ#s/ sequences in different vowel contexts: On the occurrence and the domain of sibilant assimilation. *Phonetica*, **68**, 133-160.

Pouplier, M., Hoole, P., & Scobbie, J. M. (2011). Investigating the asymmetry of English sibilant assimilation: Acoustic and EPG data. Journal of Laboratory Phonology, **2**, 1-33.

Recasens, D. (1999). Lingual coarticulation. In W. Hardcastle & N. Hewlett (Eds), *Coarticulation: Theory, Data and Techniques*. Cambridge: Cambridge University Press. Pp. 80-104.

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.

Recasens, D., Pallarès, M. D. & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America*, **102**, 544-561.

Stone, M., Faber, A., Raphael, L.J. & Shawker, T.H. (1992). Cross-sectional tongue shape and linguopalatal contact patterns in [s], /ʃ/, and [l]. *Journal of Phonetics*, **20**, 253-270.

Zharkova, N. (2008). An EPG and ultrasound study of lingual coarticulation in vowel-consonant sequences. In R. Sock, S. Fuchs & Y. Laprie (Eds), *Proceedings of the 8th International Seminar on Speech Production 2008, Strasbourg, France, 8-12 December 2008*. Pp. 241-244.

Zharkova, N., & Hewlett, N. (2009). Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and /ɑ/. *Journal of Phonetics,* **37**, 248-256.

Zharkova, N., Hewlett, N. & Hardcastle, W. J. (2011). Coarticulation as an indicator of speech motor control development in children: an ultrasound study. *Motor Control*, **15**, 118-140.

Zharkova, N., Hewlett, N. & Hardcastle, W. J. (2012). An ultrasound study of lingual coarticulation in /sV/ syllables produced by adults and typically developing children. *Journal of the International Phonetic Association,* **42**, 193-208.

Zharkova, N., Hewlett, N., Hardcastle, W. J. & Lickley, R. (accepted). Spatial and temporal lingual coarticulation and motor control in preadolescents. *Journal of Speech, Language and Hearing Research*.

Phonetic effects on the timing of gestural coordination in Modern Greek consonant clusters

Jonathan C. Yip

University of Hong Kong

Theoretical approaches to the principles governing the coordination of speech gestures differ in their assessment of the contributions of biomechanical and perceptual pressures on this coordination. Perceptually-oriented accounts postulate that, for consonant-consonant (C1-C2) sequences, gestural timing patterns arise from speakers' sensitivity to the listener's need to perceptually recover the input, whereas biomechanically-oriented accounts focus on physical factors that might constrain the relevant articulators. This dissertation contributes to current understanding of gestural coordination by examining the influences of order of place of articulation (front-to-back, back-to-front), manner of C1 (plosive, fricative), and manner of C2 (plosive, fricative, lateral) on the timing of constrictions formed by the tongue tip, tongue dorsum, and lips. If speakers produce CC sequences in order to accommodate listeners' needs, temporal separation between C1 and C2 is expected in contexts in which acoustic masking due to intergestural overlap is especially likely. If speakers' productions are instead directed by physical limitations of the vocal tract, overlap should be reduced when the gestures for C1 and C2 are not independent.

Specific instantiations of these broad hypotheses were tested in a production experiment in which eight Greek speakers' productions of initial CC sequences [pt ps pl ft kt ks kl xt] were imaged using ultrasound and video camera technologies. Degree of gestural overlap was measured in terms of temporal lag between the release of C1 constriction and the achievement of C2 constriction. Although perceptual-recoverability and biomechanical accounts made similar predictions for the effect of place order, they differed in their predictions for effects of C1 and C2 manner in the two place orders.

Results showed that, consistent with biomechanics, dorsal-coronal [kt ks kl xt] were produced with greater intergestural lag than labial-coronal [pt ps pl ft]. Consistent with perceptual recoverability, plosive-plosive [pt kt] were produced with longer lag than fricative-plosive [ft xt]. An outcome not clearly predicted by either hypothesis was that lag was longer in [pt kt] than [ps ks]. Patterns, especially for plosive-lateral [pl kl], varied across speakers. These findings revealed an interplay between physical

and perceptual—and potentially language-specific—demands on the timing of gestural coordination in speech production.

An Ultrasound Analysis of Tongue Shape in Parkinson's Disease

D. H. Whalen[1,2], Khalil Iskarous[2,3], Lucie Ménard[4], Amélie Prémont[4], Katherine M. Dawson [1]

[1]*City University of New York, [2]Haskins Laboratories, [3]University of Southern California, [4]Université du Québec à Montréal*

Parkinson's disease (PD) is a degenerative neurological disorder characterized by the degeneration of dopaminergic neurons. Dopamine is involved in the regulation of motor planning and control, and hence speech is a faculty often affected by PD. The most common diagnosis for the speech impairments in PD is hypokinetic dysarthria (Darley, Aronson, & Brown, 1969a,b). This is characterized (clinically) by slowed muscle activation, muscle rigidity, variable rate and imprecise consonant articulation.

Complex synergies of the muscles are necessary to coordinate tongue motion for linguistic purposes (Stone, Epstein, & Iskarous, 2004). It is therefore hypothesized that, as a result of the impairments above, PD subjects may produce less complex tongue shapes than control subjects during speech. In order to test this hypothesis, ten PD subjects ($\bar{x}$ - 63.8, sd - 11.3), seven older controls ($\bar{x}$ - 56.7, sd – 4.6) and ten younger controls ($\bar{x}$ - 23.1, sd - 2.2 years) were imaged using ultrasound. The subjects were native Canadian French speakers, who produced multiple repetitions of voiced stop – vowel – C stimuli.

Results will be discussed in terms of the differences in shape complexity between PD and control subjects. The suitability of different shape analysis techniques for this data will be discussed, as will future directions in using quantitative parameters to track the course of PD as it affects the speech system.

References

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12,* 246-269.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research, 12*, 462-496.

Stone, M. L., Epstein, M., & Iskarous, K. (2004). Functional segments in tongue movement. *Clinical Linguistics and Phonetics, 18,* 507–521.

Session Six

Getting away from the midline: Biomechanical devices and the role of ultrasound

Bryan Gick

University of British Columbia and Haskins Laboratories

Lingual ultrasound is most often used to image the tongue in 2-dimensional cross-section, most often following the midsagittal plane. The midsagittal section has provided generations of researchers with a simple way to reduce dimensionality in describing and modeling speech movements. This is an appealing approach, as it also helps with visualizing complex vocal tract structures while still capturing many important aspects of speech, and greatly constrains the kinds of measurements we need to make

in our speech production studies. Serrurier et al. (2012: 748) observe: "Nowadays it is largely agreed in the literature that speech movements are mainly sagittal and that information collected in the midsagittal plane is sufficient to recover the entire 3D vocal tract geometry." However, this observation makes two different points – one about how researchers might efficiently approximate the shape of the airway for acoustic modeling, and the other about how humans control their body movements. While the midsagittal plane may indeed provide useful information for an approximation of vocal tract cross-sections, it is very likely seriously flawed as a model of how dimensionality reduction is actually handled by the speech motor system – if only because the speech apparatus, with its practially unbounded degrees of freedom, has a potentially endless number of possible solutions.

An alternative model is presented based on biomechanical devices. This model handles dimensionality reduction via a set of functionally defined neuromuscular devices (drawing on, e.g., Loeb et al. 2000, Safavynia & Ting 2012, Bizzi & Cheung 2013, etc.). Ultrasound and other data as well as biomechanical simulations are presented to show how such devices may be useful in describing speech, with a focus on the role of the tongue in constrictions of the oropharyngeal isthmus (a.k.a., uvulars) following Gick et al. (2012, 2013). We note in particular that the tongue plays a highly variable role in these constrictions, and much of what is essential about the tongue's role (such as lateral bracing) occurs off midline. In the past, researchers have been constrained by tools such as x-ray and point-tracking systems that only allowed the internal structures of the vocal tract to be viewed in sagittal section or projection. However, other tools such as MRI and ultrasound have continued to perpetuate this practice, underutilizing their greater spatial flexibility. Future research will benefit from increasing the use of ultrasound to access off-midline lingual structures, and from continuing to incorporate ultrasound with other tools and modeling techniques, bearing in mind that, even for ostensibly lingual events, measuring the tongue only reveals part of the picture.

References

Bizzi, E. and Cheung, V.C.K. (2013). The neural origin of muscle synergies. *Frontiers in Computational Neuroscience*, 7: 51.

Gick, B, N. Francis, A. Klenin, E. Mizrahi and D. Tom. The velic traverse: An independent oral articulator? *Journal of the Acoustical Society of America* 133(3), EL208-EL213. 2013.

Gick, B., P. Anderson, H. Chen, C. Chiu, H. B. Kwon, I. Stavness, L. Tsou and S. Fels. (In press). Speech function of the oropharyngeal isthmus: A modeling study. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*.

Loeb, E., Giszter, S.F., Saltiel, P., Mussa-Ivaldi, F.A. and Bizzi, E. (2000). Output units of motor behavior: An experimental and modeling study. *J. Cognitive Neuroscience* 12(1):78-97.

Safavynia, S.A. and Ting, L.H. (2013). Sensorimotor feedback based on task-relevant error robustly predicts temporal recruitment and multidirectional tuning of muscle synergies. *Journal of Neurophysiology*, 109: 31–45.

An Ultrasound Study of the Acquisition of North American English /ɹ/

Lyra Magloughlin

University of Ottawa

I report an acoustic and articulatory study of North American English /ɹ/ production in typically developing English-speaking children during early- and later-stage acquisition. /ɹ/ is of interest in adult populations because it exhibits acoustic stability (e.g. low F3) despite considerable articulatory variability both within and between speakers (Delattre and Freeman, 1968; Mielke et al., 2010). Adults may use a variety of tongue shapes to achieve this articulatory target, ranging from bunched shapes with the tongue tip pointing down, to more retroflex postures with the tongue tip pointing up. In English-speaking children, /ɹ/ is often one of the last sounds to be acquired, especially in prevocalic position (McGowan et al., 2004). Tiede et al. (2011) have argued that children might attempt different vocal tract configurations during acquisition, particularly in contexts where the articulatory demands

are greater. While there is a growing body of literature on articulatory variability in adult production of /ɹ/ (e.g. Mielke et al., 2010; Westbury et al., 1998) there remains virtually no articulatory data on typically developing children's production during acquisition (although see Davidson et al., 2007, for ongoing research with atypicals).

This longitudinal study uses ultrasound imaging to investigate the articulations of four typically developing English-speaking children, aged between 3 and 6 years, during production of familiar lexical items. Children's early-stage articulations are examined and compared with their later-stage productions, and with adult variability patterns. As illustrated in Figure 1, participants' adult-like /ɹ/ productions show correspondingly low F3 values (mean: 2540 Hz), as compared with their *near*-adult-like (mean: 3172 Hz) and *non*-adult-like (mean: 3573 Hz) productions. These findings are consistent with Dalston (1975), who observed low F3 values (~2500 Hz) in the /ɹ/ productions of his child participants, in keeping with the literature on adult /ɹ/ (Delattre and Freeman, 1968; Ladefoged and Maddieson, 1996). Articulatory results show the children exhibit distinct patterns of /ɹ/ articulation, much like adults in previous studies (e.g. Mielke et al., 2010), and their development is non-deterministic, e.g. twins produce /ɹ/ with completely different (bunched vs. retroflexed) tongue shapes (Figure 2). Further, the children's exploratory behaviour mirrors adult allophony patterns, with delayed production evident in allophonic contexts that conflict with each child's dominant tongue posture. For example, one child produced exclusively bunched /ɹ/ in postvocalic and syllabic contexts (Figure 3), and was delayed in prevocalic contexts (where retroflex /ɹ/ is more common among adults) *except* prevocalically following coronals, which stands out as a context where adults often bunch, even if they retroflex elsewhere (Mielke et al., 2010). Taken together, findings provide compelling evidence to support the notion that the variability we observe with adults emerges in childhood during the exploratory period of acquisition, as suggested by Tiede et al. (2011)
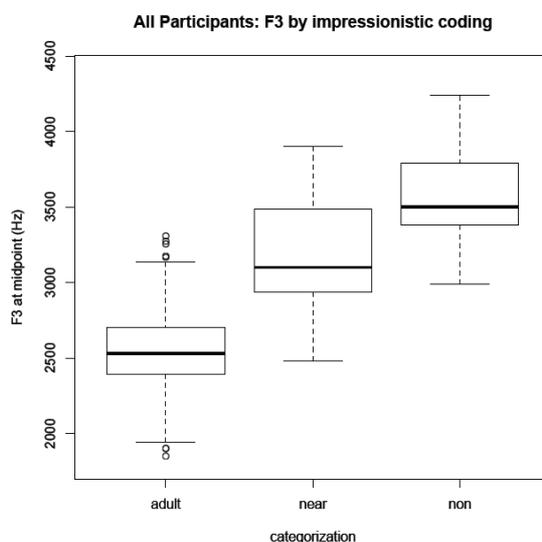
FIGURE 1. Midpoint F3 values for all participants across both Sessions, based on impressionistic categorizations as adult-like, *near*-adult-like, and *non*-adult-like.

(a)   (b)
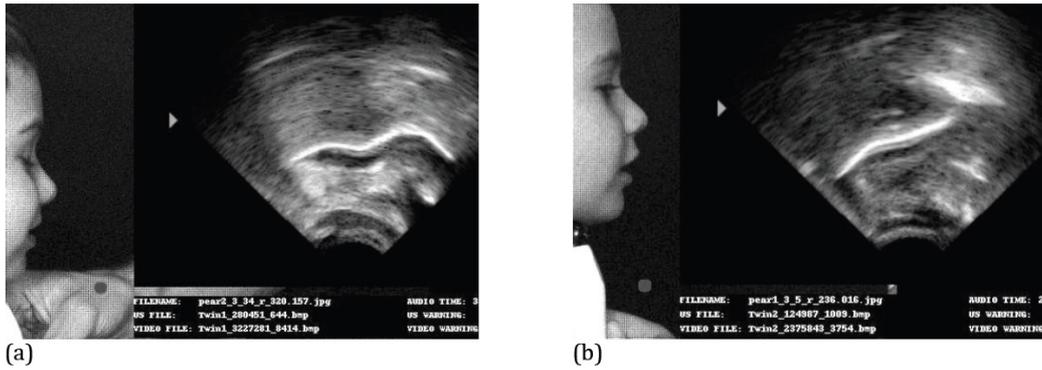
FIGURE 2. Male Twin 1, Session 2, adult-like production of postvocalic /ɹ/ in target word *pear* (a), exhibiting a bunched/tip down tongue posture, and Male Twin 2, Session 2, adult-like production of postvocalic /ɹ/ in target word *pear* (b), exhibiting a retroflex/tip up tongue posture.
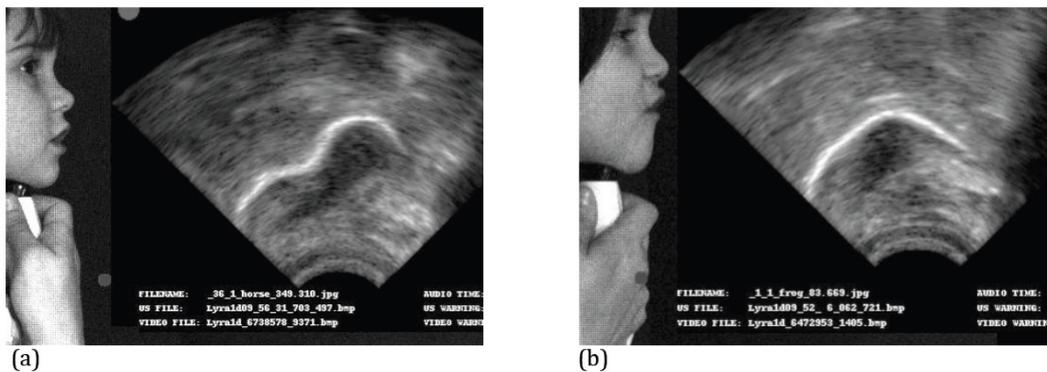

(a)   (b)

FIGURE 3. Female 1, Session 1, adult-like production of postvocalic /ɹ/ in target word *horse* (a) exhibiting concavity at the tongue dorsum and a bunched/tip down posture, and Female 1, Session 1, *non*-adult-like production of prevocalic post-labial /ɹ/ in target word *frog* (b) showing a backing of the tongue body and constriction at the velum.

References

Dalston, R. (1975). Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults. *Journal of the Acoustical Society of America,* 57:462–469.

Davidson, L., Klein, H., and Grigos, M. (2007). *Perceptual, kinematic, and ultrasound measurement of /r/ development in children with phonological delay*. Ultrafest IV, September 28-29, 2007, NYU.

Delattre, P. & Freeman, D. (1968). A dialect study of American r's by x-ray motion picture. *Linguistics*, 44:29–68.

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

McGowan, R.S., Nittrouer, S., & Manning, C.J. (2004). Development of [r] in young, Midwestern, American children. *Journal of the Acoustical Society of America*, 115(2), 871-884.

Mielke, J., Baker, A. and Archangeli, D. (2010). Variability and homogeneity in American English /ɹ/ allophony and /s/ retraction. In *Variation, Detail, & Representation*. Berlin: Mouton de Gruyter.

Tiede, M.K., Boyce, S.E., Espy-Wilson, C.Y. and Gracco. V.L. (2011). *Variability of North American English /r/ production in response to palatal perturbation*. In Speech motor control: New developments in basic and applied research, ed. Ben Maassen and Pascal van Lieshout, 53-67. Oxford: Oxford University Press.

Westbury, J., Hashi, M. and Lindstrom, M. (1998). Differences among speakers in lingual articulation for American English /ɹ/. *Speech Communication*, 26:203-226.

What happens as listeners become speakers during face-to-face conversation?

Thomas Magnuson & Chris Coey

Department of Linguistics, University of Victoria (B.C., Canada)

Approaching talk-in-interaction as an integrated system of body movement and vocalization (e.g. Bavelas et al. 2000), we collected simultaneous ultrasound and external video (Fig. 1) of eight participants' face-to-face conversations with the first author. Focusing on inter-turn intervals as well as the period immediately following the start of a turn, we sought to ascertain if there were regular patterns among body movements (including the head, trunk, and arms/hands) as well as tongue postures, swallowing, and inhalation once speech had begun. This paper presents preliminary findings for three female native English speakers.

In a given recording session, a participant watched the cartoon *Canary Row* (Warner Bros., 1950; see also McNeill 2005), after which recording began with the first author inviting the participant to narrate the cartoon as she remembered it while holding the ultrasound probe under her chin. Less-structured conversation followed the narrative elicitation. Video cameras (29.97p f.p.s.) recorded from two angles: one for gestures and body movement and the other for a profile shot consistent with the Palatron technique (Mielke et al. 2005). Ultrasound data were captured uncompressed at 29.97p f.p.s. from a Logiq-e portable ultrasound machine onto a Macintosh computer equipped to convert the VGA ultrasound signal to HDMI and synchronously record audio from a shotgun microphone. Postprocessing of the data was done using Vegas Pro 12 (Sony Creative 2012) for alignment and compilation (Fig. 2), then ELAN (MPI 2001–2013, Wittenburg et al. 2006) for frame-by-frame video segmentation and Praat (Boersma & Weenink 1992–2013) for audio segmentation.

Thus far, the data suggest that participants swallow differently when listening than when speaking. When listening, participants optionally brace the tongue against the alveolar ridge and anterior palate. During these intervals, unique swallowing events can occur up to four times before bracing is released. Swallowing also occurs during speech; however, swallowing does not occur more than once, and bracing intervals are shorter for all three speakers (Fig. 3). Three separate, single-factor ANOVAs with context (listening vs. speaking) as the independent factor and duration as the dependent variable revealed the duration difference to be significant for each speaker (Speaker 1: $F(1, 84) = 28.82$, p<.001; Speaker 2: $F(1, 32) = 19.94$, p<.001; Speaker 3: $F(1, 63) = 6.778$, p=.012). In running speech, the shorter swallowing events occur either in lieu of or immediately prior to brief inhalations, which in turn occur at fairly regular intervals at linguistically appropriate breaks (e.g. Hixon 1987).

In general, listening-related tongue bracing (and swallowing events) may co-occur with head movements (e.g. nods) and trunk movements (e.g. leaning, swiveling), as well as nasal back-channels (e.g., *"hmm?"*) but not, in these data, with manual gestures. Where nasal back-channels co-occur with head movements, the head tends to move before the onset of the acoustic signal and finish after it. This pattern of *head >> body >> speech* tends to hold for the onset of longer speech turns as well. Following the release of listening-related tongue bracing, our speakers often produced transitional utterances (e.g. *"So, uhmm…"*) as 'soft-starts' to their turns at talk. Here, too, head and body movement precedes speech. Mid-speech swallowing intervals and/or brief inhalations often occur following these transitional utterances as well. Elsewhere, once a turn at talk is fully underway, mid-speech swallowing and/or brief inhalations both take place during *hold* phases of manual gestures, while brief inhalations may also co-occur with *stroke* phases (Kendon 1980; see also Kipp 2007 for discussion of gesture phases). Manual gestures tend to begin after the onset of speech within the main portion of speakers' turns, though not exclusively.

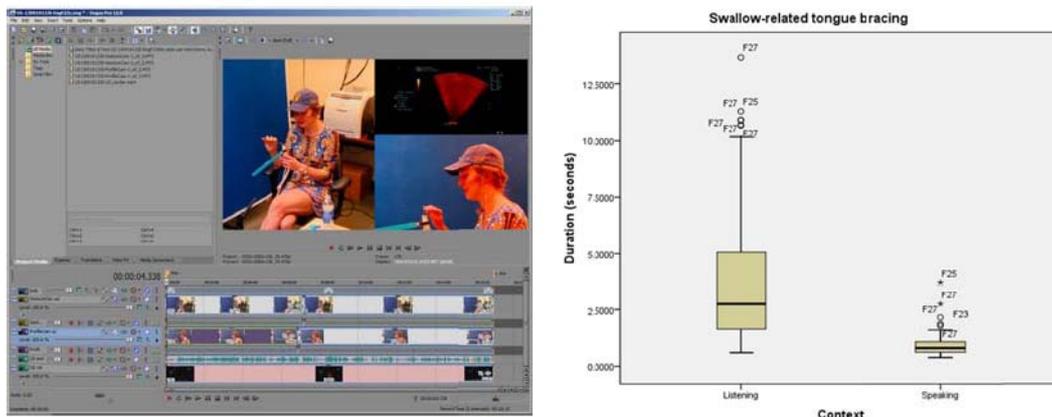**Fig. 1.** Compiled video and ultrasound data (for analysis in ELAN)

Fig. 2. Post-processing in Vegas Pro 12 Fig. 3. Swallow-related tongue bracing

References

Sony Creative Software Inc. (2012). Vegas Pro. Version 12.0 (Build 563) 64-bit.. (Computer program). htttp://www.sonycreativesoftware.com.

Bavelas, J., J. Gerwing, M. Allison, & C. Sutton (2011). Dyadic evidence for grounding with abstract deictic gestures. In: G. Stam && M. Ishino ((Eds.), *Integrating gestures: The interdisciplinary nature of gesture.* Gesture Studies 44. Philadelphia: John Benjamins, 49–660.

Boerrsma, P., & DD. Weenink (1992–2013). Praat: Doing Phonetics by Computer. Version 5.3.5. (Computer program). htttp://www.praat.org.

Hixon, T. J. (1987). *Respiratory function in speech and song.* San Diego, CA: College-Hill Press.

Kipp, M., M. Neff, & I. Albrecht (2007). An annotation scheme for conversational gestures: How to economically capture timing and form. *Language Resources & Evolution* 41, 325––339.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In: Mary Ritchie Key (Ed.), *The relationship of verbal and nonverbal communication*. The Hague: Mouton, 207––227.

McNeill, D. (2005). *Gesture and thought*. Chicago, IL: University of Chicago Press.

Mielke, J., A. Baker, D. Archangeli, & S. Racy (2005). Palatron: A technique for aligning ultrasound images of the tongue and palate. *Coyote Papers* 14, 96–107. MPII (2001–2013).

ELAN Linguistic Annotator. Version 4.6.1. (Computer program). Nijmegen, NNL: Max Planck Institute for Psycholinguistics. http://tla.mpii.nl/tools/tlaa-tools/elan/.

Warner Brothers Inc. (1950). Canary row. (Animated short film). Re-released 2001 as DVD compilation by Warner Home Video: *I Love Tweety. Volume 3*.

Wittenburg, P., H. Brugman,, A. Russel, A. Klassmann, & H. Sloetjes (2006). ELAN: A professional framework for multimodality research. In: *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LRREC),* ELRA, Paris.

Estimating Lingual Cavity Volume in Click Consonant Production from Combined Lingual Ultrasound and Palatographic Data

Amanda L. Miller

The Ohio State University

Introduction

Ultrasound data is used for linguistic description, and for speech therapy. Much of this descriptive work is based on linguistic fieldwork (Gick 2005, Miller and Finch 2011, Miller 2012a, b). Ultrasound is ideal for linguistic fieldwork given the portability of the machines, and the fact that the data collection is safe and non-invasive.

Lingual ultrasound data has been used to measure constriction locations directly. Recent research has measured constriction locations in click consonants (Miller 2012a, b) and labial-velars (Hudu et al. 2009), which are doubly articulated sounds.

Click consonants are produced using a lingual airstream mechanism. The lingual airstream in coronal click types is produced as two lingual constrictions are held in place, and the tongue body lowers to rarefy air that is trapped in the lingual cavity (Beach 1938, Ladefoged and Traill 1994, Ladefoged and Maddieson 1996, Miller, Brugman et al. 2009). Thomas-Vilakati (2008) measured the area of the lingual cavity on the palate in click consonants using electropalatography. Here, *lingual cavity volume* is estimated during the production of four contrastive coronal click types that occur in the endangered and under-described language Mangetti Dune !Xung. These four click types are listed in Table 1.

Click Type	Dental	Post-Alveolar	Lateral	Palatal
IPA Symbol	[ǀ]	[ǃ]	[ǁ]	[ǂ]

Table 1.　　　Four contrastive coronal click types in Mangetti Dune !Xung

Lingual cavity volume changes are hypothesized to be responsible for differences in loudness of the click bursts of the 4 different click types.

Methods

Lingual cavity volume was estimated at three points in time that characterize three stages during the production of all four click types by 4 speakers of Mangetti Dune !Xung The CHAUSA method was employed to collect data at 114 fps, which yields an image of almost the entire tongue, every 8.77 ms. A single word for each click type was produced 12 times by 4 speakers of Mangetti Dune !Xung, in Namibia.

The tongue edge was traced from lingual ultrasound images at three critical stages in the production of the 4 click types. The first lingual trace is at the beginning of the overlap phase, when both constrictions are initially formed. The second trace is the midpoint between the first trace and the third trace. The third trace is the moment just prior to the anterior release of the click. The cavity volume is hypothesized to increase from stage 1 to stage 3, which is expected to have the greatest cavity volume.

The methodology for measuring the lingual cavity volume at these three points in time is provided in (1) – (8):

1) The top of the lingual cavity was estimated from stone palate casts that were made for each of the speakers in Mangetti Dune, Namibia, using the methodology described in Ladefoged

(1957). The stone palates were digitized using a 3D Poldhemus digitizer, and were represented by 41 points, using the methodology described in Ferreiro et al. (1998) . Figure 1 provides a map of the placement of the 41 points on the palate.

2) 2D palates were traced from ultrasound movies of each speaker swallowing following the methodology in Epstein and Stone (2005).

3) Each speaker's 2D palate from ultrasound, and their 3D digitized palate from the stone palate cast, was aligned spatially by aligning the two highest vertical points on the digital palates.

4) The y and z displacement numbers between the 2D and 3D palate representations were calculated from this vertical peak, and all 41 points of the 3D palate were transformed on the y and z axes to the 2D palate space.

5) The tongue length was estimated from the average tongue length computed over all 12 repetitions of each of the clicks that were recorded for each speaker.

6) The cavity width measurements were made from 5 points in a single palatogram of each click made for each speaker's productions (anterior width, point between the anterior width and the widest point, the widest point, point midway between the widest width and the narrowest width, and the narrowest width at the other end).

7) The cavity volume was estimated by calculating the volume of 0.010 cm trapezoids that were placed along the lingual cavity, and summing the volumes of all of the trapezoids.

8) Changes in estimated cavity volume were calculated between trace 3 and trace 1, in order to estimate the total change in volume that occurs between the beginning of the overlap phase, when the lingual cavity is first formed (trace 1), and the end of the rarefaction phase, which is just prior to the anterior constriction release (trace 3).

(a) **(b)**
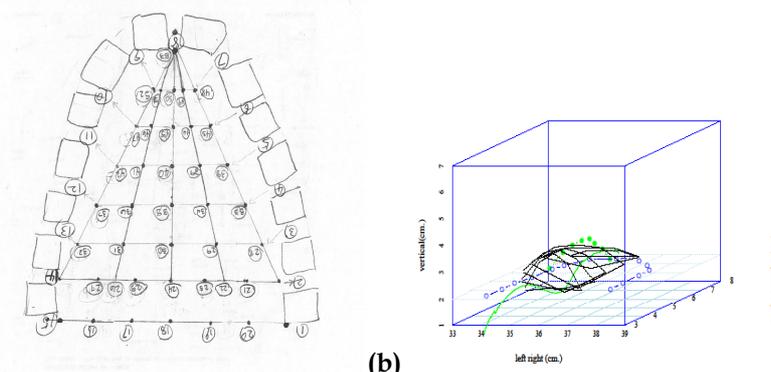
Figure 1. **(a)** Map of 41 points scanned to digitize stone palate casts; **(b)** 3D plot of the lingual cavity in the Alveolar click, [!], at the end of the overlap phase (after rarefaction takes place, stage 3)

Results

Table 2 provides the results of the change in lingual cavity volume found between trace 1 and trace 3, for 3 out of 4 coronal click types in Mangetti Dune !Xung.

Click Type	Lingual Cavity Volume at end of overlap phase	Lingual Cavity volume at begin of overlap phase	Change in Lingual Cavity Volume During Rarefaction
Dental [ǀ]	5.54 cm^3	4.59 cm^3	**0.95 cm^3**
Alveolar [!]	7.74 cm^3	5.08 cm^3	**1.93 cm^3**
Lateral [ǁ]	5.26 cm^3	4.65 cm^3	**0.61 cm^3**

Table 2. Lingual cavity volume change from begin of overlap phase in clicks to the moment before the tongue tip release (end of overlap phase) for all four coronal click types in Mangetti Dune !Xung

Discussion

The change in cavity volume between trace 1 and trace 3 is related largely to the loudness of the click bursts. Traill (1997), Miller-Ockhuizen (2003), and Nakagawa (2006) have shown that the [!] click is the loudest, followed by the [ǀǀ] click, followed by the [ǀǀ] and [ǂ] clicks. The change in lingual cavity volume estimates provided in Table 2 above show the largest volume change for the alveolar [!] click

type, which is in keeping with this click type being the loudest click. However, the cavity volume changes show that the dental click has a greater change in volume during rarefaction than the dental click type. This is opposite what we see with respect to the relative loudness of these click types. Both the dental and lateral click types are noisy (Sands 1990; Johnson 1993; Traill 1997; Miller and Shah 2007), which Thomas-Vilakati (2010) shows is due to an additional noise source as the tongue pulls away from the front teeth, and side teeth respectively. I surmise that the louder noise burst in the lateral click, than is found in the dental click, is due to the larger contact area found between the tongue and the sides of the teeth.

References

Articulate Instruments Ltd. (2008). Ultrasound stabilization headset user's manual, Revision 1.3. Edinburgh: Articulate Instruments Ltd.

Ferreiro, V. F., Sforza, C. Schmitz, J. H. and Colombo, A. (1998). Quantitative Description of the Morphology of the Human Palate by a Mathematical Equation. The Cleft Palate-Craniofacial Journal, 35(5): 396-401.

Hudu, F., Miller, A. and Pulleyblank, D. 2009. .Proceedings of Conference on Language Documentation & Linguistics Theory 2. P. K. Austin, O. Bond, M. Charette, D. Nathan & P. Sells, Eds., SOAS: London, . 153 - 163.

IPA. (2006). The International Phonetic Alphabet (revised to 2005) [chart]. *Journal of the International Phonetic Association* **36(1)**, 135.

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

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

Mielke, J., Baker, A., Archangeli, D. & Racy, S. (2005). Palatron: a technique for aligning ultrasound images of the tongue and palate. In Siddiqi, D. & Tucker, B. V. (Eds.), *Coyote Papers 14*, 97-108.

Miller, A. (2012a). C-V Co-articulation in Consonants with Multiple Lingual Constrictions. Under Review. *Journal of Phonetics.*

Miller, A. (2012b). The Representation of Mangetti Dune !Xung Clicks: Rarefaction Gestures and Release Gestures. MS. The Ohio State University.

Miller, A. and Finch, K. (2011). Corrected high frame rate anchored ultrasound with software alignment. April 2011, *Journal of Speech, Language and Hearing Research*, 471-486.

Miller, A. and Shah, S. 2009. The Acoustics of Mangetti Dune !Xung Clicks. In M. Uther, R. Moore & S. Cox, Eds., *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*. Causal Productions: Brighton, U.K., 2283-2286.

Miller-Ockhuizen, A. 2003. *The Phonetics and Phonology of Gutturals: A Case study from Juǀʼhoansi,* Outstanding Dissertations in Linguistics Series (Laurence Horn, Ed.), New York: Routledge.

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

Stone, M. (2005). A Guide to Analyzing Tongue Motion from Ultrasound Images. *Clinical Linguistics and Phonetics 19, 6-7*, 455-502.

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

Traill, A. (1997). Linguistic phonetic features for clicks: articulatory, acoustic and perceptual evidence. (Herbert, R. K., Ed.), *African linguistics at the crossroads: papers from Kwaluseni.* Köln: Rüdiger Köppe Verlag.

/s/ in Italian-Tyrolean bilingual speakers

Lorenzo Spreafico

Free University of Bozen-Bolzano

This paper presents an exploratory UTI-based study on articulatory patterns for /s/C(C) clusters in Italian. More specifically, it focuses on word-initial and word-internal /s/-stop-(liquid) sequences as produced by four Italian/Tyrolean simultaneous and sequential bilingual speakers. The main aim of

this research is to establish if the articulatory patterns for /s/ in Italian vary according to the nature and degree of bilingualism of the speaker.

In this respect, the first step is to elicit the articulatory patterns for /s/ and assess whether they are influenced by the nature of the consonant(s) that follow(s). The intra-speaker comparison for tongue profiles based on the SS ANOVA approach (Davidson 2006) shows that the consonant triple /p t k/ exert a small coarticulatory effects on /s/ midpoint (Fig. 1a, b, c).

The second step is to investigate whether the tongue configurations for /s/ differ systematically according to the position in the word. The comparison between /s/ tongue contours in word-initial and word-internal position displays that the shape of the tongue as well as its location and orientation (Iskarous 2005) are affected (Fig. 2a, b, c).

The third step is to examine if /s/ as uttered in word-initial /spr/, /str/ and /skr/ clusters are more retracted than in /sp/, /st/ and /sk/ sequences (Mielke et al. 2010). Smoothing spline estimates and Bayesian confidence intervals applied to the tongue profiles of the four bilinguals do not confirm this hypothesis (Fig. 3a, b, c).

The final step is to correlate all data on articulatory patterns and tongue configuration in simultaneous and sequential bilingual speakers and comment on the relationship between degree of bilingualism and variation of articulatory patters.

All articulatory data presented in the paper are captured using an Ultrasonix SonixTablet system coupled to a transducer operating at 5-9 MHz and linked to the Articulate Instruments AAA software for data acquisition and analysis at a temporal resolution of 122/166Hz.
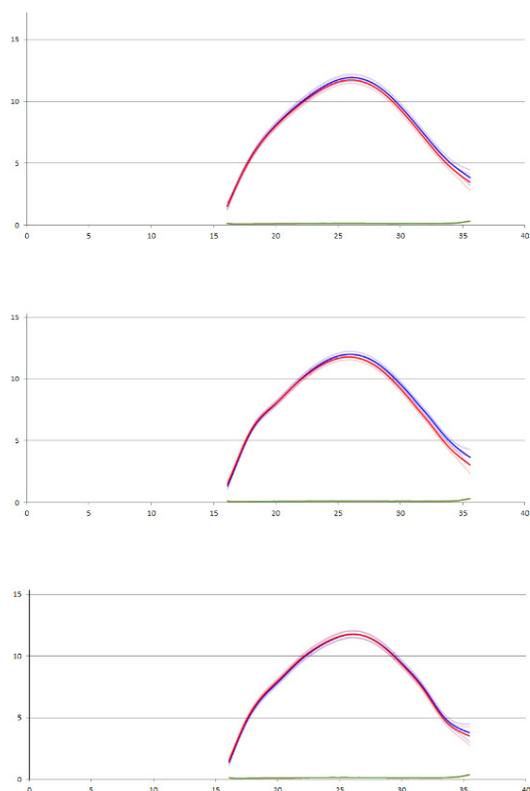
Figure. 1 Smoothing spline estimate and 95% Bayesian confidence interval for comparison of the of mean curves for /s/ in "scanno", "spazzo", "stato".
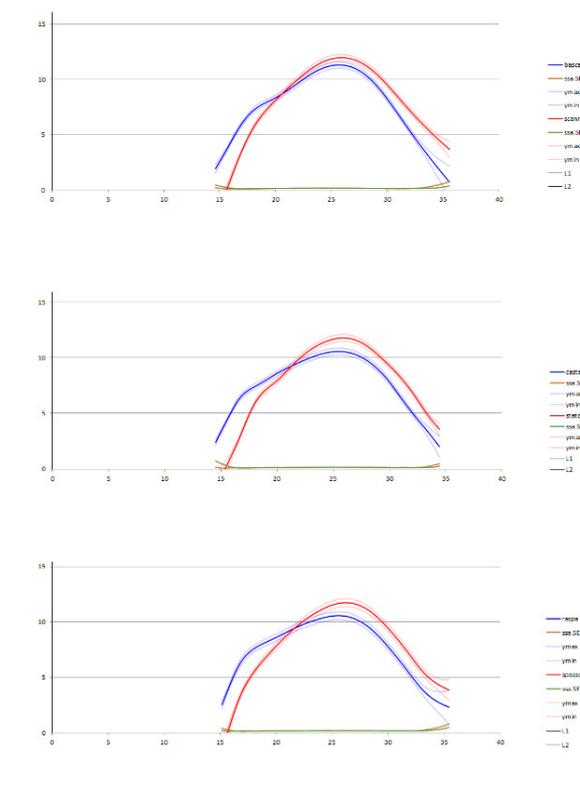
Figure. 2 Smoothing spline estimate and 95% Bayesian confidence interval for comparison the mean curves for /s/ in "basca" and "scanno "; "casta" and "stato"; "raspa" and "spazzo".
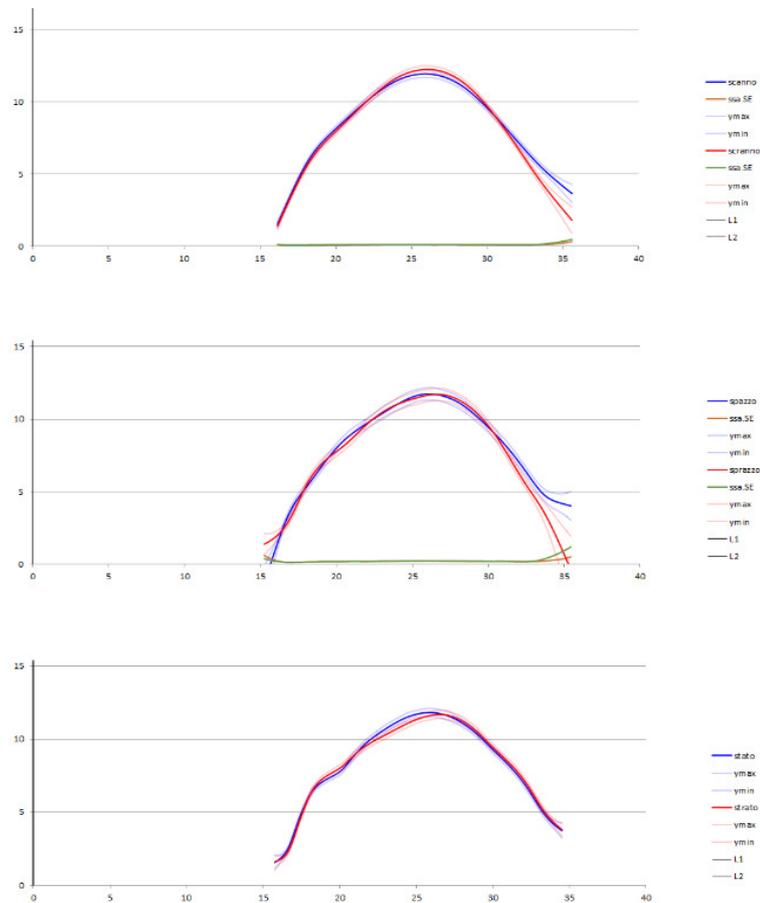
Figure. 3 Smoothing spline estimate and 95% Bayesian confidence interval for comparison of the mean curves for /s/ in "scranno" and "scanno"; "strato" and "stato"; "sprazzo" and "spazzo".

References

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

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

Mielke J. et al. (2010) *Variability and homogeneity in American English /r/ allophony and /s/ retraction*. In Fougeron C. et al (eds.), Variation, Detail, and Representation. Berlin: Mouton de Gruyter: 699-719.

The Japanese unrounded back vowel /ɯ/ is in fact rounded central/front [ʉ - ʏ]

Akitsugu Nogita   **Noriko Yamane**   **Sonya Bird**

University of Victoria   University of British Columbia   University of Victoria

This study reports on an ultrasound and video recording investigation of the so-called unrounded back vowel /ɯ/ in Standard Japanese (SJ henceforth), as in /ɯɕiɾo/ 'back', based on our hypothesis that

/ɯ/ is underlyingly a rounded central vowel with lip protrusion and the more appropriate phonemic IPA symbol is /ʉ/. As it turns out, the tongue position of SJ /ɯ/ was closer to that of the front vowel /e/ than the back vowel /o/ by 6 of the 7 native SJ speaking participants, and all the 7 participants actively rounded their lips; at least 4 of them also showed clear lip protrusion.

Background and research questions

In terms of *lip rounding*, linguists commonly treat SJ /ɯ/ as unrounded. In fact, the prescriptive /ɯ/ is supposed to be with flat lips rather than rounded lips (Akiyama, 2009). In contrast, Ito, Kang and Kenstowicz (2006) state that SJ /ɯ/ is produced with vertical lip compression, or narrowing of the lips but without lip protrusion. As for *backness*, some linguists treat both SJ /ɯ/ and /a/ as central vowels as opposed to the back /o/. In fact, some acoustic studies, such as Hisagi, Nishi and Strange (2008), show that the second formant (F2) of /ɯ/ is very close to that of /a/, rather than /o/, but F2 of /ɯ/ seems slightly lower than that of /a/. If /ɯ/ is in fact central, the commonly used phonemic or broad transcription /ɯ/ is inappropriate. Thus, this study questions whether native speakers of SJ actively unround their lips as seen in the Korean /ɨ/ as in /hankɨl/ 'Hangul' and actively back their tongue, as the symbol /ɯ/ indicates. More specifically, does /ɯ/ lack lip protrusion as Ito et al. mention? Is the tongue posture of /ɯ/ articulatorily closer to back vowels or front vowels?

To answer these questions, we had 7 linguistically naïve native SJ speakers pronounce the five short vowels /i, e, a, o, ɯ/. Each phoneme was pronounced in isolation in order to have the participants emphasize underlying features of each phoneme. The tongue and lip movements were video-recorded with an ultrasound machine and a video camera. Each phoneme was pronounced 12 times. The tongue shapes of each phoneme and rest position were traced in EdgeTrak, and analyzed in R.

Results and discussion

The results indicate that /ɯ/ is actively *rounded* by all the participants, as opposed to a rest position with no active rounding. At least 4 of the 7 participants showed clear lip protrusion as shown in Figure 1 below.

Figure 1. Lip protrusion in the "unrounded" /ɯ/ by a female Standard Japanese speaker in her 20's and a male speaker in his 30's.

The analysis of backness in the ultrasound images suggests that /ɯ/ is more front than /a/ for all the speakers as shown in Figure 2. More interestingly, the tongue position of /ɯ/ is closer to that of the front vowel /e/ than that of the back vowel /o/ by 6 out of the 7 participants (see Figure 2) and the rest of the speaker's /ɯ/ is located almost in the middle between /o/ and /e/. As well, /ɯ/ is closer to the rest position than other vowels. Moreover, among 5 out of the 7 participants, the front part of the tongue of /ɯ/ patterns together with /i, e/ and the rest position rather than with /o, a/ as shown in Figure 3. These findings suggest that /ɯ/ is *central* or even slightly front, rather than back.

To answer our questions, therefore, 1) /ɯ/ is underlyingly *rounded* and it can involve lip protrusion, and 2) /ɯ/ is more front than the so-called central /a/, and typically closer to /e/ than /o/. Thus, /ɯ/ is *central* but more towards the front than towards the back. Therefore, a possible IPA for /ɯ/ in isolation is [ʉ], or even [ʏ] in some cases (but not as front as [y]). The reason why F2 of /ɯ/ is slightly lower than that of /a/ may be caused by lip rounding.

Conclusion

To conclude, /ɯ/ is not the appropriate symbol for the phonemic transcription of Standard Japanese. We propose the rounded central vowel /ʉ/: e.g. /ɯɕiɾo/ → /ʉɕiɾo/.

References:

Akiyama, K. (2009). *Sensê seminar: Hatsuon hassê*. NHK CTI Nihongo Centre.

Hisagi, M., Nishi, K., & Strange, W. (2008). Acoustic properties of Japanese and English vowels: Effects of phonetic and prosodic context. In M. Endo-Hudson, S.-A. Jun, P. Sells, P. M. Clancy, S. Iwasaki, and S. Sung-Ock, eds., *Japanese/Korean Linguistics 13*. Stanford: CSLI.

Ito, C., Kang, Y., & Kenstowicz, M. (2006). The adaptation of Japanese loanwords into Korean. *MIT Working Papers in Linguistics* 52: 65-104.
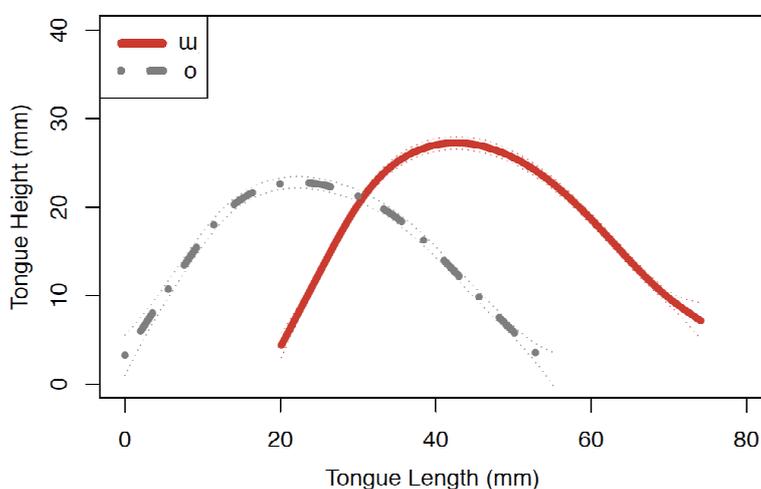
M20A /ɯ/ vs. /o/
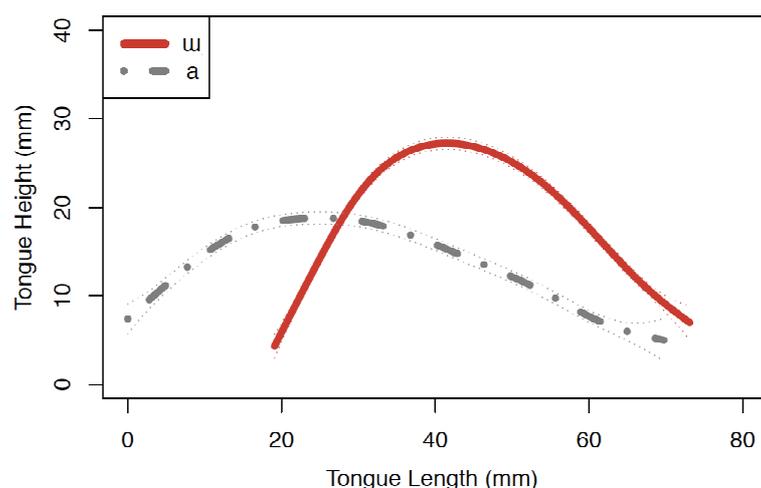
M20A /ɯ/ vs. /a/

M20A /ɯ/ vs. /e/
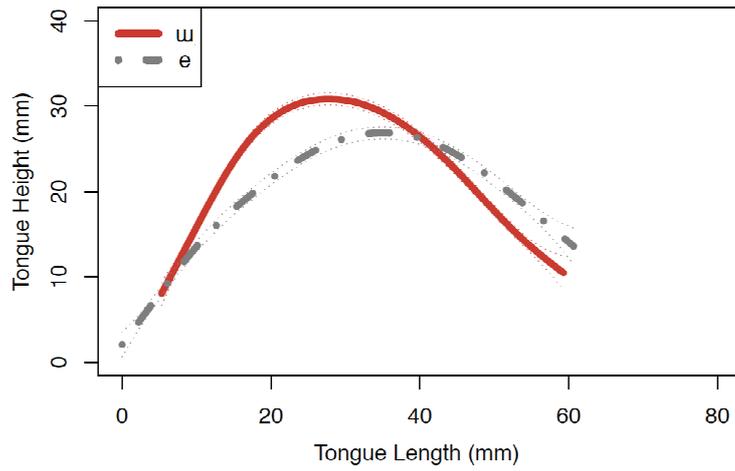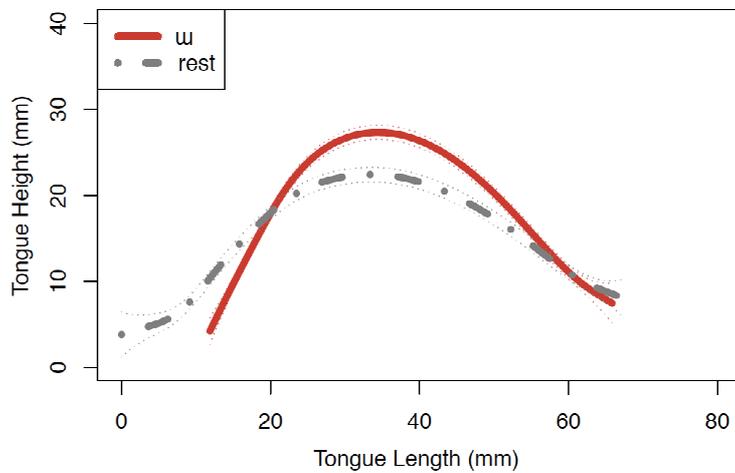
M20A /ɯ/ vs. the rest position

Figure 2. The tongue positions of /ɯ/ vs. /o/, /ɯ/ vs. /a/, /ɯ/ vs. /e/, and /ɯ/ vs. the rest position by a male speaker in his 20's (labeled as M20A).
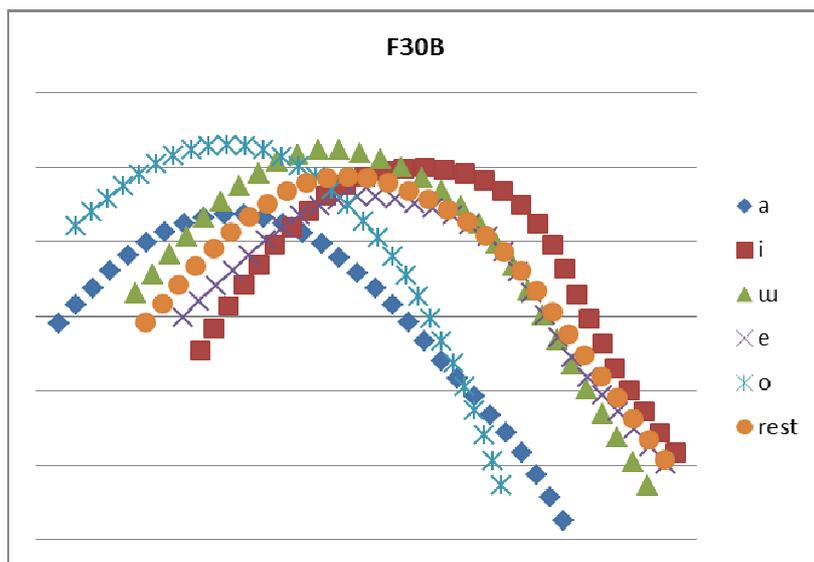
Figure 3. The average tongue contours of /a, i, ɯ, e, o/ and the rest position by a female speaker in her 30's (labeled as F30B).

42

The Low Mandible Maneuver

Angelika Nair[1], Gareth Nair[1] and Maureen Stone[2]

1.Drew University, Music Dept.   2. University of Maryland, School of Dentistry

Classical singers of international rank appear to optimize the resonance of their sound by allowing the condyle to descend in the temporomandibular joint (TMJ). This action drops the posterior mandible and creates an increase in oral cavity resonance space. Furthermore, the maneuver allows a drop of the larynx, an event that has considerable effect on the available resonance space of the pharynx. This combined maneuver can more than double the resonance space. Its use also requires a considerable retraining of tongue configuration due to the need to achieve the phonetic point-of-articulation (POA) with the mandible platform dropped.

While this critical technique has been occasionally referred to in the pedagogical literature, it has not been scientifically investigated because true internationally-ranked singers have not been studied in the laboratory. In order to fill this void, our study was designed in two phases; 1) a pilot study to explore both the physiology of the technique as well as the efficacy of the use of ultrasound (with corroborating MRI, imagery and acoustic analysis) as a research tool, 2) (still to be executed) taking a portable ultrasound machine to the studios, dressing rooms and homes of internationally ranked singers to explore the use of the low-mandible maneuver in the world of today's internationally-ranked classical singers.

This presentation will reveal the investigative techniques and results of the pilot study. In this phase of the research, we teamed up with the Medical University of Graz, Austria. We investigated the techniques of five regionally-ranked singers utilizing both the ultrasound and MRI. The MRI images were valuable in confirming the efficacy of the ultrasound imagery in the study.

We were able to gather a wealth of imagery that both confirms the use of the Low-Mandible Maneuver and provides the first scientific insight into its physiology. At the same time, acoustic analysis can be applied to the sung signals to document the result of the increase in resonance space the maneuver affords.

The Low Mandible Maneuver has rather significant ramifications for resonance production by enabling a concomitantly lowered larynx and increased resonance space in the pharyngeal and oral cavities. Its use also has a rather significant effect on the tongue shapes required for all sung phonemes. This presentation will employ videos, acoustic analysis (spectrography) as well as stop-action imagery to illustrate the Low Mandible Maneuver.

Lateral tongue bracing in Japanese and English

Ian Wilson, Julian Villegas, Terumasa Doi

Center for Language Research, University of Aizu

Coronal ultrasound imaging will be used to compare the degree of lateral tongue bracing that occurs in English with that occurring in Japanese. The speech of Japanese speakers of English as a second language will be examined to test the hypothesis that those who brace more (as is thought to be normal for English native speakers) have pronunciation that is perceived to be closer to native-like.

Arabic Emphatics and Gutturals

Majed Al Solami

University of Toronto

The articulation of Arabic guttural class and emphatics and their effects on adjacent vowels have been the subject of many studies in the literature of Arabic phonetics and phonology (Al-Ani, 1970; Delattre, 1971; Ali & Daniloff, 1972; Ghazeli, 1977; Alwan, 1989; McCarthy, 1994; Zawaydeh, 1999; inter alia). These studies show different and to some extent inconsistent mechanisms for the articulation of Arabic laryngeal, pharyngeal, uvular and emphatic sounds. Also, they reported different effects that the tongue retraction triggers in neighboring vowels. Furthermore, previous studies have examined the articulation of Arabic gutturals and emphatics using a form of endoscopy, which does not show the tongue root (TR) and tongue dorsum (TD) during the articulation of these sounds.

The current study uses ultrasound imaging of the tongue coupled with acoustic signal to examine the articulations involved in guttural and emphatic sounds in three Arabic dialects, Egyptian, Saudi and Palestinian Arabic. This investigation attempts to answer two questions: (i) what are the physiological mechanisms involved in the production of Arabic guttural sounds, both inherently retracted /ʕ, ħ, ʁ, χ, q/ and secondarily retracted /d̪, t̪, s̪, ð̪/ in terms of TR and TD retraction? and (ii) what are the coarticulatory effects of these sounds on neighboring vowels?

Articulatory results indicate that these sounds are produced with different TR and TD retraction mechanisms. Pharyngeals are articulated with TR retraction and statistically do not involve significant TD retraction. Uvulars and emphatics, on the other hand, show TR and TD retraction with uvulars showing inconsistent TR retraction. This suggests that pharyngeals are articulated with TR retraction, while uvulars and emphatics retract TR as a result of TD retraction. Laryngeals do not show any significant tongue retraction.

The acoustic experiment, using nonsense word speech samples from nine male subjects, three subjects from each dialect, examined the coarticulatory impact of Arabic emphatic and guttural sounds on formant frequencies of adjacent vowels. Results show that pharyngeals are associated with high F1 transitions while uvulars and emphatics are associated with low F2 transitions. Also, low F2 transition was stronger in emphatics than in uvulars. Laryngeals are not associated with any specific coarticulatory effects on adjacent vowels.

Apart from the benefit of articulatory and acoustic descriptions of Arabic gutturals and emphatics in terms of tongue shapes and vowel retraction effects, the findings of this study raise some questions regarding the phonological representation of Arabic gutturals and emphatics.

References

Al-Ani, S. (1970). Arabic phonology; an acoustical and physiological investigation. The Hague: Mouton.

Ali, L.H., Daniloff, R.E. (1972). A contrastive cinefluorographic investigation of the articulation of emphatic-non-emphatic cognate consonants. Studia Linguistica 26, 81-105.

Alwan, A. (1989). Perceptual cues for place of articulation for the voiced pharyngeal and uvular consonants. Journal of the Acoustical Society of America, 86, 549-556.

Bin-Muqbil, M. (2006). Phonetic and phonological aspects of Arabic emphatics and gutturals. PhD dissertation, University of Wisconsin-Maddison.

Delattre, P. (1971). Pharyngeal features in the consonants of Arabic, German, Spanish, French, and American English. Phonetica, 23, 129-155.

Ghazeli, S. (1977). Back Consonants and Backing Coarticulation in Arabic. PhD, Univ. Texas at Austin.

McCarthy, J. (1994). The phonetics and phonology of Semitic pharyngeals. In P. Keating (ed.), Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III.Cambridge: Cambridge University Press.

Zawaydeh, B. (1999). The phonetics and phonology of gutturals in Arabic, Ph.D. Dissertation, Indiana University.

Velar coarticulation and tongue twisters in people who stutter 2: Disfluent utterances

Stefan A. Frisch, Alissa Belmont, & Nathan D. Maxfield

University of South Florida

Introduction

This paper will present data from across the lifespan for people who stutter. The task involved repeating velarvowel consonant combinations in a carrier sentence (e.g. Say a <u>cope</u> again, following Wodzinski & Frisch 2005 ASA presentation) followed by a set of tongue twisters designed to induce velaralveolar errors (e.g. top cap cop tab, following Pouplier & Goldstein 2005). Disfluent productions by people who stutter were analyzed to determine similarities and differences with productions of people with typical speech production.

Analysis

In this phase of the project, disfluencies were examined qualitatively by observing ultrasound video and timelocked broadband spectrograms using the Articulate Assistant Advanced software (Articulate Instruments, 2012). In select cases, tongue traces were created to compare tongue postures during instances of disfluency with fluent productions of the same targets.

Participants

30 speakers who stutter have been recorded, 10 each from three age groups (children 812, young adults 1839, older adults 5065). Analysis is ongoing as of the abstract deadline.

Results

To date, disfluencies were observed in all 10 young adults, one older adult, and 6 children who stutter. The disfluencies can be generally classified into three types:

1. Typical speech errors – These disfluencies are interruptions or hesitations in the vicinity of a speech error. At a qualitative level, these difluencies appear to be no different than disfluencies that have been observed in people with typical speech production (Frisch, previous ultrafests). These were observed in most participants who had disfluencies on target sounds during the experiment.
2. Moments of oral stuttering – Apparently stuttered oral articulation appeared to occur in one of two ways:
    a. Participants locked or perseverated on a target articulation (cf. Sheehan & Voas 1954). These disfluencies were observed in one young adult and two children
    b. Participants rapidly repeated or fluttered on a target articulation (cf. Denny & Smith 1992). These disfluencies were observed in two young adults, one older adult, and one child.
3. Moments of laryngeal stuttering – Apparently stuttered laryngeal articulation was observed (cf. Thürmer, Thumfart, & Kittel 1983). These moments of disfluency seemed to involve severe laryngeal tension. Retraction and lowering of the tongue body was observed during these moments consistent with infrahyoid muscle contraction affecting tongue position via the hyoglossus and mylohyoid. These disfluencies were observed in two younger adults and one child.

Retroflex versus bunched /r/ in ultrasound intervention: One shape does not fit all?

Tara McAllister Byun[1] and Elaine R. Hitchcock[2]

1. NYU   2. Montclair State University

Errors in speech sound production can have a negative impact on children's development across academic, social, and psycho-emotional domains. While most children go on to attain age-appropriate speech production, an estimated 30% of English-speaking children with a history of speech sound disorder continue to misproduce rhotic sounds at 9 years of age, and 9% still show these errors from 12-18 years and beyond (Lewis & Shriberg, 1994). Clinicians have called for more effective treatment methods for residual speech errors, especially for the phoneme /r/ (RSE-/r/). Several descriptive studies have suggested that ultrasound biofeedback can succeed in eliminating /r/ errors that have not responded to other forms of intervention (Adler-Bock et al., 2007; Modha et al., 2008; Bernhardt et al., 2008). There is a need to follow up on these promising results with systematic, controlled studies in order to build an evidence base to support the use of ultrasound biofeedback for RSE. This two-part study aimed to use single-subject experimental methods to systematically measure response to ultrasound biofeedback therapy.

Participants in Study 1 were four monolingual English-speaking children, ages 6;1-10;3, who produced /r/ at the word level with less than 30% accuracy but otherwise showed normal speech, language, and hearing abilities. In a multiple-baseline across-subjects design, participants began in a no-treatment baseline phase whose duration increased by one session for each successive client enrolled. Participants then received biofeedback intervention in twice-weekly 30-minute sessions for 8 weeks. In treatment sessions, participants produced 60 trials of /r/ at the syllable level while viewing an ultrasound image generated with an Interson SeeMore USB probe. Both consonantal and vocalic variants of /r/ were targeted in each session. Participants received verbal cues encouraging them to match their tongue shape to a static ultrasound image representing a bunched /r/ tongue shape.

To measure progress, all /r/ trials elicited during baseline, within-treatment, and maintenance probes were isolated and presented in a randomized, de-identified fashion for rating by clinician listeners. Interrater agreement exceeded 88%. Individual trajectories of accuracy over time are depicted in Figure 1. Effect sizes measuring the change from baseline to maintenance period ranged from -1.4 to 2.2 (mean = 0.6), revealing generalization gains of disappointingly small magnitude. Even within the treatment setting, visual inspection of blinded ratings revealed that only one participant ("Gabby", age 10;3) made substantial, sustained gains. A second study was thus undertaken to determine whether a modified treatment protocol could enhance the gains produced by ultrasound biofeedback. The modification of the study was inspired by an observation about the participant who improved during Study 1: her progress occurred after she happened to produce a retroflex tongue shape which yielded a perceptually accurate /r/. The clinician then reinforced this tongue shape instead of using the standard cues for bunched /r/, and "Gabby's" accuracy increased steadily, reaching a maximum of 48% within the treatment setting. Thus, it was speculated that outcomes could be improved by using individualized tongue shape targets.

Study 2 enrolled a new group of four children, ages 7;8-15;8, who met the same criteria outlined for Study 1. The same multiple-baseline across-subjects design was used, and treatment followed the same protocol, with one exception: prior to the first treatment session, an ultrasound recording of a contextual /r/ probe was created for each child. Based on the shape of the tongue during the child's closest /r/ approximations, a candidate tongue shape was selected from among the MRI images collected by Tiede et al. (2004). A retroflex tongue shape was selected as a primary target for one participant and a bunched tip-up shape was targeted for two participants; for the fourth participant, retroflex and bunched shapes were judged to be equally facilitative. Outcomes, again measured with blinded clinician ratings, differed dramatically from Study 1. Across individuals and treatment targets (consonantal vs vocalic /r/), effect sizes ranged from 1.0 to 16.7, with a mean of 7.3. Individual trajectories are depicted in Figure 2.
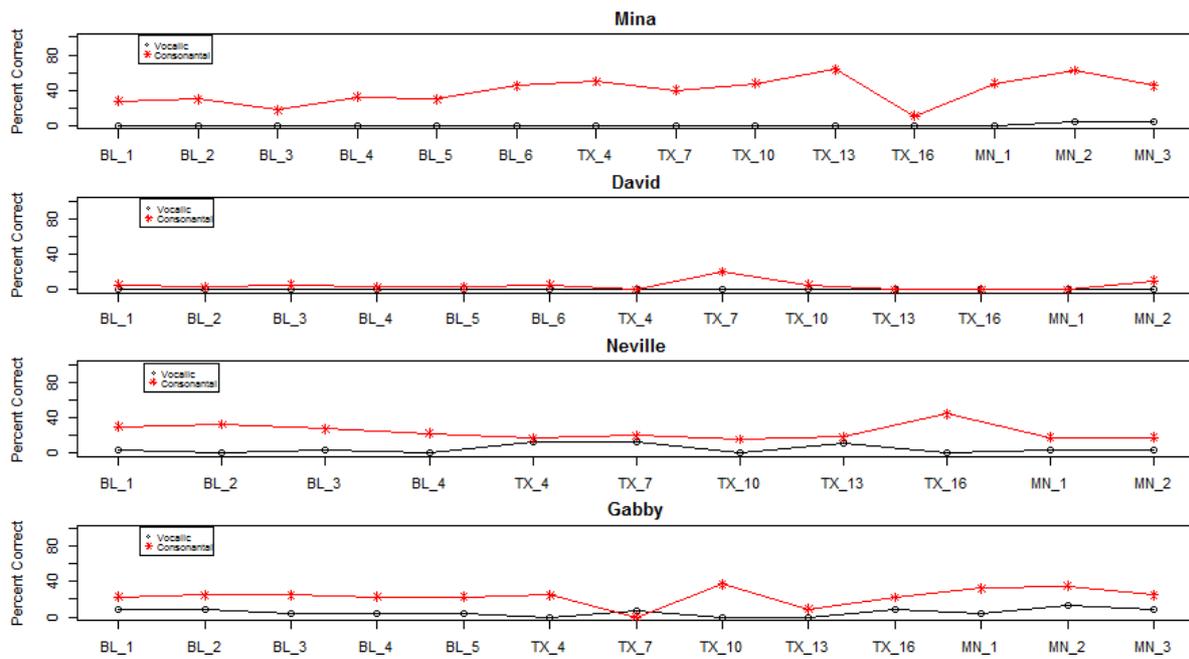
Figure 1. Percent correct on generalization probe measures, Cohort 1
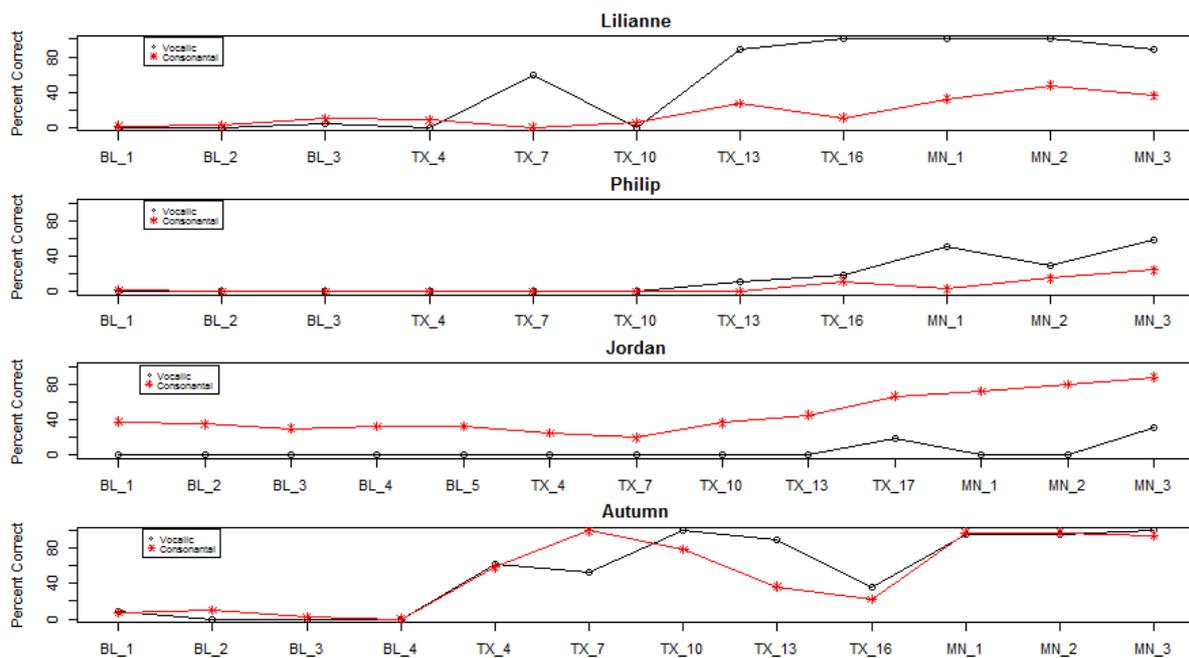

Figure 2. Percent correct on generalization probe measures, Cohort 2

These two studies do not constitute a controlled comparison of ultrasound treatment with and without the option to select individualized tongue shapes. Other differences, such as the older mean age of participants or the greater experience of the treating clinician in Study 2, may have contributed to the observed difference in outcomes. This research must be followed up in a more systematic fashion before strong conclusions can be drawn. Still, our results offer a strong suggestion that ultrasound biofeedback treatment for RSE-/r/ should include opportunities to explore different tongue shapes in order to find the best fit for the vocal tract characteristics of the individual child.

References

Adler-Bock, M., Bernhardt, B. M., Gick, B., & Bacsfalvi, P. (2007). The use of ultrasound in remediation of North American English /r/ in 2 adolescents. *American Journal of Speech-Language Pathology*, 16, 128-139.

Bernhardt, B., et al. (2008). Ultrasound as visual feedback in speech habilitation: Exploring consultative use in rural British Columbia, Canada. *Clinical Linguistics and Phonetics*, 22, 149-162.

Modha, G., et al. (2008). Case study using ultrasound to treat /r/. *International Journal of Language & Communication Disorders*, 43, 323-329.

Lewis, B. A., & Shriberg, L. D. (1994). Life span interrelationships among speech, prosody-voice, and nontraditional phonological measures. Miniseminar presented at the ASHA convention, New Orleans, LA.

Tiede, M. K., Boyce, S. E., Holland, C. K., & Choe, K. A. (2004). A new taxonomy of American English /r/ using MRI and ultrasound. *Journal of the Acoustical Society of America*, 115, 2633-2634.

Ultrasound biofeedback therapy for –r-distortions: Is it worth the effort?

Tim Bressmann (1,2), Irina Zhylich (1), Gajanan V. Kulkarni (2)

(1) Department of Speech-Language Pathology, University of Toronto

(2) *Faculty of Dentistry. University of Toronto*

r-distortions are often persistent and difficult to treat. Adler-Boch et al. (2007) and Modha et al. (2008) have recently reported that visual feedback using ultrasound can be a valuable adjunct for the speech therapy of r-distortions. In the present study, 8 patients received individual speech therapy for their r-distortions. Of these, 5 received conventional therapy, supplemented by 10 minutes of ultrasound biofeedback per session. The ultrasound biofeedback was presented on a computer screen with a visual overlay of a cartoon vocal tract. The other 3 patients received conventional speech therapy without ultrasound biofeedback. The treatment outcomes of the two groups were compared with regards to perceptual, acoustic and motor outcomes. The results were variable, and progress was only modest for most participants. The use of the ultrasound biofeedback did not appear to make a critical difference in the treatment outcomes in the present study. Possible implication for participant selection and treatment dosage will be discussed.

Can seeing help you hear? Acoustic vs. Ultrasound information in the diagnosis of speech sound disorders associated with cleft palate

Zoe Roxburgh, James, M. Scobbie, Joanne Cleland and Sara Wood

CASL, Queen Margaret University

Many studies investigating speech characteristics in cleft palate (CP) have been based on perceptual analysis. In most cases, phonetic transcription is deemed to be gold standard in diagnosing speech sound disorders associated with cleft palate (Sell 2005). Due to the heterogeneity and complexity of CP, atypical and compensatory articulations may not be identified through phonetic transcription alone. For speech and language therapists using phonetic transcription, there is a risk of these unusual patterns being unidentified, with possible misdiagnosis and subsequent inappropriate intervention.

Technological advances allow for a more detailed assessment of cleft palate speech, using instrumental techniques. One such tool, Electropalatography (EPG), indirectly shows the effects of the cleft using a standardised palate for those with typical, or in the case of CP, atypical, palate shapes and sizes. Howard (2004) investigated the compensatory articulations in adolescents with cleft palate using perceptual and EPG information. It was concluded that information provided by EPG allowed insight into a range of atypical patterns which were unidentified during phonetic transcription alone.

A more recent tool, ultrasound tongue imaging (UTI), also shows the indirect effect of the cleft on speech. As UTI shows the surface of the tongue from nearly the tip to the root, it is particularly useful for looking at backing (Stone 2005), a common compensatory articulation in CP speech (Peterson-Falzone et al. 2010). Gibbon and Wolters (2005) and Bressmann, Radovanovic, Kulkarni, Klaiman and Fisher (2011) have explored the compensatory articulations in CP speech using UTI and have concluded that it has the potential to become a useful tool for investigating cleft palate speech. In addition, Zharkova (2013) has identified ultrasound as a potentially useful tool for the assessment of speech in speakers with CP. However its clinical applications remain to be tested, particularly as a biofeedback tool in intervention.

The current study tests and compares the clinical application of articulatory animations and UTI for speech sound disorders in children with a repaired cleft palate. One key aspect of this study is to identify whether ultrasound confirms phonetic transcriptions and if it provides any additional information on the compensatory articulations in CP speech described in the literature.

This study uses a single-subject multiple-baseline design. Participants are three males, ages 6;3, 6;7 and 9;2, with a secondary speech sound disorder as a result of repaired CP.

Participants received two assessment sessions, in which the phonology subtest of the Diagnostic Evaluation of Articulation and Phonology (DEAP) (Dodd et al. 2002) and phoneme-specific wordlists were administered and recorded using an Ultrasonix® SonixRP machine remotely controlled via Ethernet from a PC running Articulate Assistant Advanced™ software (AAA) (Articulate Instruments 2010). A headset was used to stabilise the probe, to ensure accurate measurements were gathered. To ensure that headset movement was accounted for, a video from a headset-mounted micro-camera was used, which also captured lip data. A headset-mounted microphone was also used to record audio data.

A narrow transcription of speech data was carried out by two phonetically trained listeners, using symbols from the International Phonetic Alphabet chart and the Extended International Alphabet chart. Lip data was taken into consideration for perceptual analysis. Ultrasound data was analysed using AAA software version 2.14 (Articulate Instruments 2010).

Phonetic transcriptions and ultrasound data were analysed in detail for speaker 2, a 6 year old male with repaired submucous cleft. Phonetic transcriptions of the DEAP and phoneme-specific wordlists revealed cleft type characteristics such as backing of alveolar and velar to glottal place of articulation, velopharyngeal friction and possible double articulations. A preliminary analysis of ultrasound data confirmed phonetic transcriptions of speaker 2. It also revealed covert errors such as double articulations of alveolar-glottal stops. Lip data revealed additional silent labial articulations. A detailed analysis of findings from speaker 2 will be presented.

References:

Articulate Assistant Advanced Ultrasound Module User Manual, 2010. Revision 212, Articulate Instruments Ltd.

Bressmann, T., Radovanovic, B., Kulkarni, G. V., Klaiman, P., and Fisher, D. 2011. An ultrasonographic investigation of cleft-type compensatory articulations of voiceless velar stops. *Clinical Linguistics & Phonetics*. Vol. 25(11-12) pp.1028-1033.

Dodd, B., Hua, Z., Crosbie, S., Holm, A., and Ozanne, A. 2002. *Diagnostic Evaluation of Articulation and Phonology.* London: The Psychological Corporation.

Gibbon, F.E. and Wolters, M. 2005. A new application of ultrasound to image tongue behaviour in cleft palate speech. Poster presentation at the Craniofacial Society of Great Britain and Ireland Annual Scientific Conference, Swansea, UK (13-15 April 2005).

Howard, S. 2004. Compensatory articulatory behaviours in adolescents with cleft palate: comparing the perceptual and instrumental evidence. *Clinical Linguistics & Phonetics*. Vol 18. pp. 313-340.

Peterson-Falzone, S.J., Hardin-Jones, M.A., and Karnell, M.P. 2010. *Cleft Palate Speech*. 4th ed. United States: Mosby, Elsevier.

Sell, D. 2005. Issues in perceptual speech analysis in cleft palate and related disorders: a review. *International Journal of Language and Communication Disorders*. Vol 40(2). pp.103-121.

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

Zharkova, N. 2013. Using ultrasound to quantify tongue shape and movement characteristics. *The Cleft Palate-Craniofacial Journal*. Vol 50. pp 76-81.

Poster 1:

Capturing, Analyzing, and Transmitting Intangible Cultural Heritage with the i-Treasures Project

A. Jaumard-Hakoun1,2, S. K. Al Kork[1],2, M. Adda-Decker3, A. Amelot3, L. Buchman3, G. Dreyfus2 , T. Fux3, P. Roussel2, C. Pillot-Loiseau3, M. Stone4, B. Denby1,2

1 Université Pierre et Marie Curie, Paris, France
2 Signal Processing and Machine Learning Lab, ESPCI-ParisTech, Paris, France
3Phonetics and Phonology Lab, CNRS UMR 7018, University Paris 3 Sorbonne Nouvelle
4 Vocal Tract Visualization Lab, University of Maryland Dental School, Baltimore, USA

Project overview

The i-Treasures project, which officially began on 1 February 2013, is a 12-partner FP7 project that proposes to use multi-sensor technology to capture, preserve, and transmit four types of intangible cultural heritage, referred to as 'use cases': rare traditional songs, rare dance interactions, traditional craftsmanship and contemporary music composition. Methodologies used will include body and gesture recognition, vocal tract modeling, speech processing and electroencephalography (EEG). The "Rare traditional songs" use case, which will be the focus of our work, targets Corsican "cantu in paghjella", Sardinian "canto a tenore", Mt. Athos Greek byzantine hymns and the recent "Human beat box" styles. The final objective of the "Rare traditional songs" use case is to capture vocal tract movements with sufficient accuracy to drive a real-time 2D or 3D avatar of the vocal tract, which will in turn play a crucial role in the transmitting of captured invisible cultural heritage to future generations.

Acquisition system

To study vocal tract dynamics during singing performances, artists will be instrumented with a non-intrusive acquisition helmet (see figure 1) containing an ultrasound probe, a camera and a microphone. Additional sensors will complete the acquisition system: a piezoelectric accelerometer mounted on the nose, an Electroglottograph (EGG) necklace and a breathing belt. The ultrasound probe will be used to study tongue movements, articulation and lingual contour [1]. The camera will help us detecting lips and jaw movements, lips aperture and protrusion. The microphone provides an acoustic reference of sound production and enables spectral analysis. Nasality will be measured thanks to the accelerometer while the EGG will provide information concerning glottis dynamics. The kind of breathing as well as breathing rhythm (amplitude, frequency) will be explored through the breathing belt. Information concerning vocal quality will be extracted from speech, EGG and accelerometer signals [2, 3]. Both ultrasound probe and camera provide grayscale pictures of size respectively 320 x 240 and 640 x 480 pixels acquired at 60 fps. Signals from the other sensors are mono-dimensional and sampled at 44100 Hz. The proposed acquisition system is developed using a Real Time Modular Application System

(RTMaps) [4] and provides real-time data displaying and recording, which can be either local or on a network.
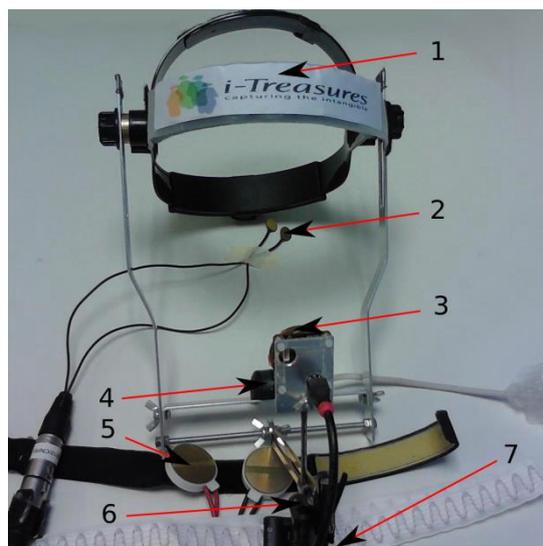
1. Multi-sensor helmet
2. Nose-mounted accelerometer
3. Camera
4. Ultrasound probe
5. Electroglottograph
6. Microphone
7. Breathing belt

Figure 1: Acquisition sensors

Vocal tract modeling

Capturing and modeling intangible cultural heritage is a challenging task that will require appropriate sets of salient descriptors derived from sensor data. Currently, data-driven "deep learning" or DL architectures are being investigated as a means of extracting articulatorily pertinent vocal tract information from ultrasound of the tongue and lips images. The DL approach, which has been applied to pattern or phoneme recognition, is one of the most efficient unsupervised learning methods. A typical DL architecture consists of a neural network with a large number of hidden layers, with each additional layer corresponding to a higher level of abstraction, a hierarchy which imitates the layered structure of the brain's visual and auditory cortexes. Pre-training such a network, in an unsupervised way instead of initializing it randomly, provides both optimization and regularization of the network and reduces training error [5].

Preliminary data processing

Such DL architectures applied on ultrasound images may be relevant to extracting salient descriptors from articulatory data and modeling vocal tract movements during singing performance. We aim at extracting the contour of the tongue from ultrasound images in a speaker-independent manner. If we train a network on a large database of several singers' recordings, the intuition is that our network will be able to extract descriptors regardless the configuration of the mouth, provided the network is sufficiently trained. Another reason for using DL is that once the network is optimized for a specified task, it is easy to use it to perform this task without any human supervision. Processing one image only requires one single pass through the network, which allows real-time processing. Assuming we have a contour for a subset of our ultrasound data (manually or automatically extracted), our network is trained on a representative learning database composed for each example of a reduced image of ultrasound data and a contour image. Our input contour images are obtained thanks to an automatic tongue contour extraction algorithm: for each columns of each image, several pixels are selected as contour candidates. Decision is made according to the shape of the previous image and the distance between the candidate and the last point detected as contour. In our work, as shown in figure 2, we first reduce the dimensions of input images (both ultrasound and contour images are reduced to 30x33 pixels) so that the number of neurons in the input layer of our neural network is not too large.

Extracting tongue contour with deep networks

We used an autoencoder structure (encoder followed by a decoder, respectively the lower and the upper part of the network, see figure 2) so that the network is trained to reconstruct the input from the descriptors of its hidden layers, which contains an abstract representation of the inputs. The construction of this autoencoder is made through greedy layer-wise Restricted Boltzmann Machines (RBM) training [6]. A RBM is a neural network made of a visible unit and a hidden unit defined by its

probability distribution (Bernoulli). The propagation rule is given by joint probabilities and weights are updated according to a learning rule. A deep network is made of stacked and sequentially-trained RBM, so that the hidden layer of each RBM is the visible unit of the following RBM. Hidden layers contain a compressed representation of input data that seems sufficient to reconstruct a contour from sensor-only input data using "translational" RBM defined in [7]. The idea is to train a network on both ultrasound and contour images as input to learn a representation of these inputs and their relationship. Given these shared features, the decoder stage is able to reconstruct both images. Then if a new encoder can be trained on ultrasound data only to produce hidden features, identical to those produced by the network with concatenated ultrasound and contour data, it is possible to reconstruct labels from hidden features using the decoder of the original network (see figure 3).
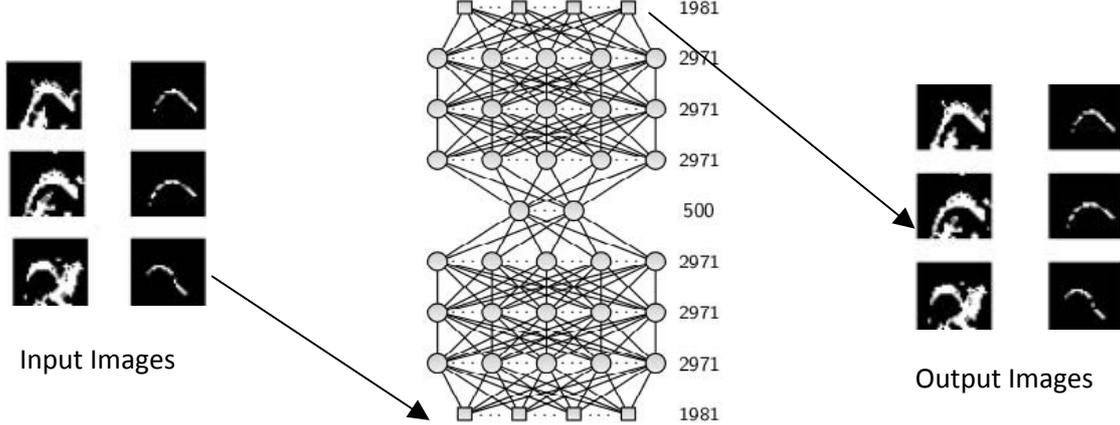
Figure 2. Deep autoencoder training..

The first layer of our network is composed of one neuron per pixel of the ultrasound image and one neuron per pixel of the contour image plus one for bias, that is 1981 neurons (see figure 2). Our network is composed of 7 hidden layers of various lengths (figure 2 and 3). Once our deep autoencoder is trained to reconstruct input images (figure 2), we use our translational network on new ultrasound images to reconstruct both ultrasound and contour images (figure 3). Despite the low quality of input images, we are able to convert output images into a set of coordinates that represents the contour of the tongue. In figure 4, red points represent the output of the translational network after conversion into pixel coordinates, which combines contour image thresholding, isolated points removal and contour interpolation. However, the quality of reconstructed contour can be very poor if the input image is too noisy and quite sensitive to artifacts. Improving input image quality can be considered as one option, although it would dramatically increase computing time.
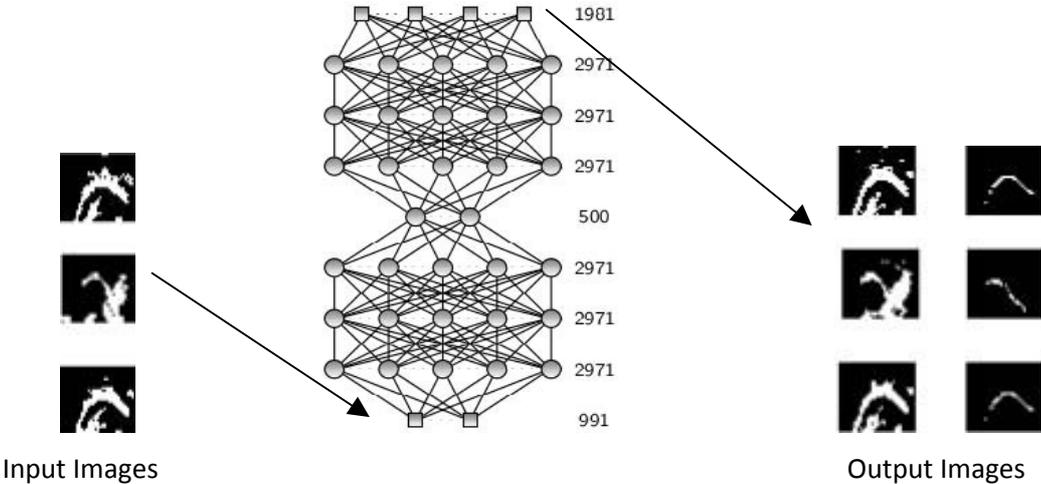
Figure 3. Translational network use.
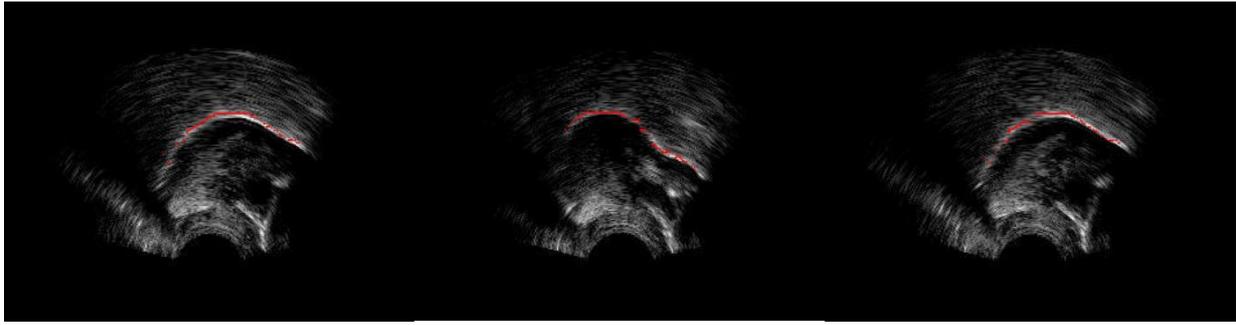
Figure 4. Examples of tongue contour extraction results.

Conclusions and future work

Using a DL approach is a useful way to extract salient descriptors from ultrasound images. The advantage of our method is the use of automatically extracted contours as inputs to train our neural network so that the network is able to learn tongue contour extraction. We will extend this method to lip images processing. In addition, since our current sensor processing algorithms are offline tests, our next steps will concern the integration of a sensor output processing block to our current real-time module.

References

[1] M. Stone, *A Guide to Analyzing Tongue Motion from Ultrasound Images,* Clinical Linguistics and Phonetics, 19(6-7), 455-502, 2005.

[2] N. Henrich, C. D'Alessandro, B. Doval, M. Castellengo, *Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency,* The Journal of the Acoustical Society of America 117(3), 1417-1430, 2005.

[3] C. D'Alessandro, *Method in empirical prosody research*, chapter "Voice Source Parameters and Prosodic Analysis", 63–87, 2006.

[4] INTEMPORA S.A., 2011. [Online]. Available: http://www.intempora.com/.

[5] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent and S. Bengio, *Why Does Unsupervised Pre-training Help Deep Learning ?,* J. Mach. Learn. Res., 2010.

[6] G. Hinton, S. Osindero, and Y. The, *A fast learning algorithm for deep belief nets*, Neural Computation, 18(7),1527–1554, 2006.

[7] I. Fasel and J. Berry, *Deep Belief Networks for Real-Time Extraction of Tongue Contours from Ultrasound During Speech*, ICPR, 1493-1496. IEEE, 2010.

Poster 3:

Statistical analysis of tongue shape in ultrasound video sequences: tongue tracking and population analysis

Sarah Ghrenassia[1], Catherine Laporte[1], and Lucie Ménard[2]

[1] *Laboratoire de traitement de l'information en santé, École de technologie supérieure, Canada*
[2] *Laboratoire de phonétique, Université du Québec à Montréal, Canada*

Introduction and background

In this study, we developed and evaluated a set of automated and semi-automated image processing tools for the statistical analysis of tongue shape, as seen in the midsagittal view on 2D ultrasound video sequences. It is known that tongue shape and motion, though highly non-rigid, are also highly constrained: most speech patterns can be explained by a few degrees of freedom. These, in turn, are

well captured by so-called active shape models (ASMs) (Cootes et al., 1995) based on Principal Component Analysis, which we use to derive the principal modes of variation in tongue shape. We fitted an ASM to a training database of 1978 tongue contours extracted from ultrasound video sequences of two adult subjects uttering [VpV] and [VtV] sequences in French (where V is one of the ten French oral vowels) shows that 95% of the shape variations are explained by three principal components. We then successfully applied the ASM framework to two different tasks related to ultrasound-based speech analysis: (1) tongue tracking and (2) phonological characterization.

Improvement of tongue tracking using ASMs

We first applied the ASM framework to the task of tongue tracking in ultrasound video sequences. Specifically, we built upon the well-known active contour approach of Li et al. (2005), which is implemented for instance within their widely used *EdgeTrak* software, and used our ASM to constrain the active contour to agree with plausible tongue contour shapes, as previously suggested by Hamarneh and Gustavsson (2000). As illustrated in Fig. 1, this can reduce the occurrence of tracking errors (and the time spent manually correcting them), even for utterances not included in the training database. Improvements were particularly marked in the presence of rapidly changing tongue curvature, as in the utterance [uku]. These findings corroborate previous work by Roussos et al. (2010) who also developed a shape model for ultrasound tongue tracking using PCA, but trained the model on X-ray images. In contrast, we showed that ultrasound images are sufficient to obtain a useful shape model, making the approach less invasive.
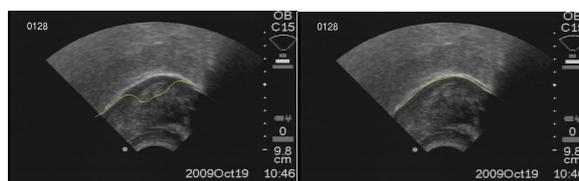
Figure. 1 – Left: ultrasound image frame with tracking error (active contour model only). Right: same image frame without tracking error (ASM-constrained active contour model).

Using ASMs for phonological population analysis

We then used our ASM to describe tongue shapes in ultrasound video sequences of two adults and two 4-year-old children uttering [VpV] and [VtV] sequences in French. We fitted the ASM to tongue contours extracted from each frame in the video sequences and plotted each shape model parameter as a function of time. We coarsely aligned the curves corresponding to different repetitions of individual utterances by the same subject using the shortest utterance as a reference. Fig. 2 shows the first shape model parameter (the weight of the first principal component) as a function of time for each subject and repetition of the utterance [yty] after alignment. While the fit of the ASM was very good for adults and children alike, plotting the evolution of the parameters corresponding to each principal component showed interesting differences between adult and child populations, with adult parameters showing far more repeatability over similar utterances than child sequences, agreeing with the discussion by Smith (2010). These preliminary results demonstrate the potential of ultrasound-based ASMs as tools to reveal and analyze significant tongue shape or motion differences between populations.
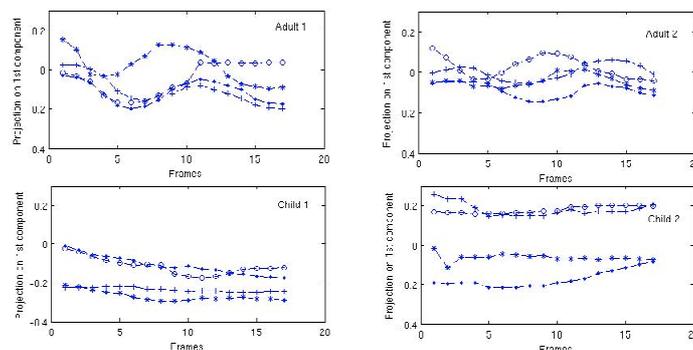

Figure. 2 – Evolution of the first ASM parameter over time during [yty] utterances by two adults (top row) and two children (bottom row).

Conclusions

ASMs are a potentially powerful tool both for the automated/semi-automated extraction of tongue contours from ultrasound images and for the temporal analysis of tongue shape variability among and between different phonological populations.

References

Cootes, Timothy F., et al. "Active shape models-their training and application." *Computer vision and image understanding* 61.1 (1995): 38-59.

Hamarneh, G. and T. Gustavsson. "Combining snakes and active shape models for segmenting the human left ventricle in echocardiographic images." *Computers in Cardiology* 27 (2000): 115-118.

Li, Min, Chandra Kambhamettu, and Maureen Stone. "Automatic contour tracking in ultrasound images." *Clinical linguistics & phonetics* 19.6-7 (2005): 545-554.

Roussos, Anastasios, Athanassios Katsamanis, and Petros Maragos. "Tongue tracking in ultrasound images with active appearance models." *16th IEEE International Conference on Image Processing,* 2009.

Smith, Anne. "Development of Neural Control of Orofacial Movements for Speech." In Handbook of Phonetic Sciences. W. Hardcastle, J. Laver, & F. Gibbon (Eds). Oxford: Blackwell, 2010

Poster 5:

Ultrasound images in the new 'iPA Phonetics' App

Christopher Coey[1], John H. Esling[1], Scott R. Moisik[2]

[1]University of Victoria, [2]Max Planck Institute for Psycholinguistics

The phonetic chart in an App

iPA Phonetics is an application that illustrates the sounds and articulations of an expanded version of the IPA chart. The App gives users of Apple iOS mobile electronic devices the ability to access and compare and, through matching 'games,' to test their knowledge of phonetic symbols and sounds together with their visual production correlates, including video of the oral vocal tract, video of the laryngeal vocal tract, and ultrasound of the laryngeal vocal tract. The App is entirely 'self-contained' in that once downloaded it requires no internet connection to use.

The chart format follows the design of the elaborated chart of speech sounds of the IPA published in Esling's chapter on notation in: Hardcastle, Laver & Gibbon (Eds.), *The Handbook of Phonetic Sciences,* 2nd ed. Oxford: Wiley-Blackwell (2010). The consonant chart contains 14 columns for place of articulation and 13 rows for manner of articulation. There are 4 major regions delimited in the chart – separating labial, tongue-front, tongue-back, and laryngeal sounds. Diacritic qualifications of sound quality are included in the extra symbols themselves that appear in the chart, some of which derive from the ExtIPA (Duckworth et al. 1990, Ball 1991). There are 223 consonant symbols/sounds in the chart. A vowel chart is also included, which follows the format of the 2005 IPA chart. There are 28 symbols/sounds in the vowel chart, with two styles of production: long and short. Video/audio recordings associating the production of a phonetic sound with each symbol were captured using the KayPENTAX 9105 70°-angle rigid oral laryngoscope fitted with a 35mm lens to capture oral articulations and a 28mm wide-angle lens to capture laryngeal articulations (Esling 1996, 1999) and connected through the Panasonic GP-US522 camera to video/audio/ultrasound channel synchronizing software. Camera angles begin with the scope outside of the mouth, progressing further inside of the mouth for tongue-back recordings, and over the back of the tongue to view laryngeal activity. The video/audio recordings are accessed in the App by clicking on consonant symbols or vowel symbols. A

diagram of the Laryngeal Articulator Model of the vocal tract (Esling 2005) is also included, elaborated with a set of clickable voice quality labels, representing sound quality associated with various parts of the vocal tract, and accompanied by illustrative recordings.

Ultrasound illustrations in the App

Ultrasound images of consonant sound production have been included together with the laryngoscopic video/audio for the last two columns of the consonant chart (the laryngeal region: Pharyngeal/Epiglottal and Glottal). This articulatory region of the vocal tract has always been elusive to visualize, and ultrasound techniques applied to this region are rare. Ultrasound images were captured using a GE portable LOGIQe R5.0.1 system with an 8C-RS (convex) probe to image supraglottal laryngeal involvement (e.g. for Glottal stop) and with an 12L-RS straight-line probe at a relatively shallow 2-4 cm depth on the neck and about 2-4 cm of the vertical dimension for clear resolution of laryngeal structures to image larynx height changes (e.g. for Pharyngeal/Epiglottals). This is a novel laryngeal technique that differs from the approach usually taken in oral lingual ultrasound data capture (Moisik 2013).
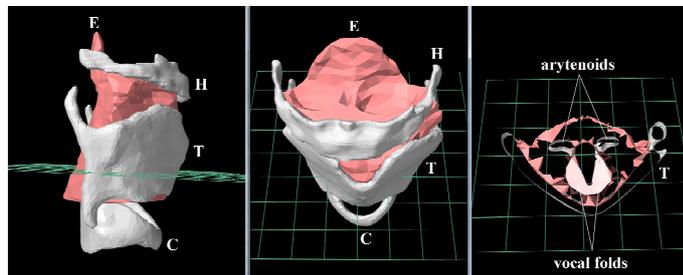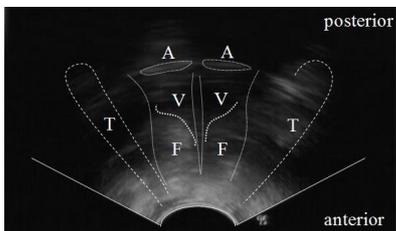
Figure. 1. Axial convex ultrasound.

Figure. 2. Axial section of the larynx (in ArtiSynth).

Key to the labels in the diagrams and ultrasound views:

A = arytenoid cartilage, AE = aryepiglottic fold, C = cricoid cartilage, E = epiglottis,

F = ventricular fold, H = hyoid bone, L = vocal ligament, T = thyroid cartilage, V = vocal fold
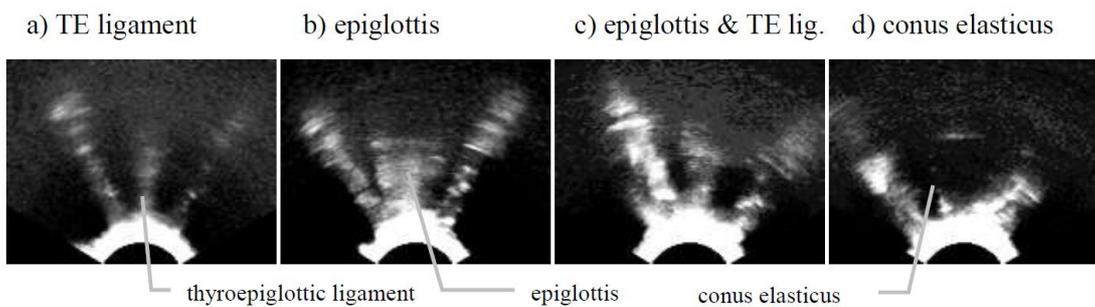

a) TE ligament   b) epiglottis   c) epiglottis & TE lig.   d) conus elasticus

thyroepiglottic ligament   epiglottis   conus elasticus

Figure. 3. Axial convex ultrasound, sub- and supraglottal views.

a) relaxed TA   b) contracted TA

arytenoid cartilage

thyroid cartilage

ventricular fold (left)
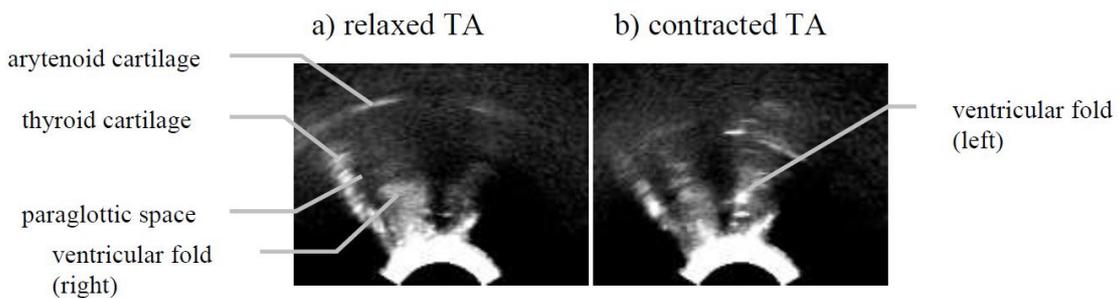
paraglottic space
ventricular fold (right)

Figure 4. Axial convex ultrasound, ventricular folds.

Three ultrasound viewing options for laryngeal sound production are included in the App. The first perspective was captured with the 8C-RS (convex) probe in an axial (horizontal) orientation at the front of the larynx. The second perspective was captured with the 8C-RS (convex) probe in a vertical orientation at the front of the larynx. The third perspective was captured with the 12L-RS (flat) probe in a quasi-vertical orientation towards the side of the larynx. The ultrasound viewing option can be selected with a toggle switch.

The axial (horizontal) convex ultrasound view is shown in Figure 1. This orientation corresponds to the physical structures of the larynx in Figure 2, which have been modelled in the ArtiSynth environment (Moisik & Gick 2013). The horizontal section in Figure 2 is at the level of the vocal folds. The anatomical sections in the models are shown in parallel with the corresponding labelled ultrasound images to illustrate what can be seen using laryngeal ultrasound. Both the labelled ultrasound diagrams and laryngeal model diagrams appear in the beta version of the iPA Phonetics App, assembled onto one 'INFO' page, viewable by pressing the info button next to the US-video ANGLE buttons. Figures 3 and 4 illustrate various epilaryngeal-level (see Figure 5) structures visible in the axial convex ultrasound view. Figure 6 depicts the near-coronal or paracoronal view (see Figure 7) obtained from vertical orientation of the convex probe. Note that the red dotted line in Figure 7 delineates the vocal and ventricular folds and roughly corresponds to the white dotted line in Figure 6. Finally, Figures 8 and 9 demonstrate the vertical imaging plane captured using the flat probe held upright. With this probe, the penetration depth tends to be weaker, but the structural resolution of the superficial layers of tissue and the thyroid lamina are much more crisply imaged. Figure 9 depicts a series of images obtained by sliding the probe along the thyroid lamina posterior to anterior. This series provides an impression of the three-dimensional geometry of the lamina and the surrounding muscular and connective tissues.
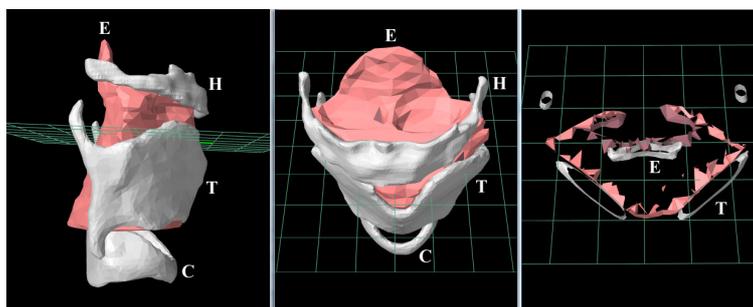

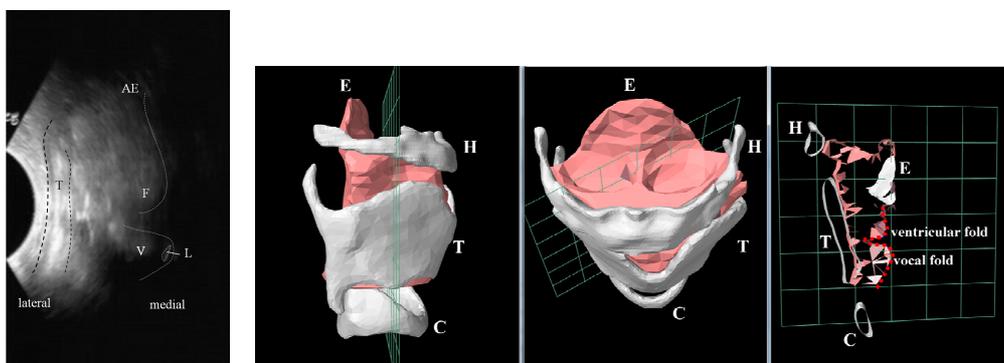**Figure. 5.** Axial section of the larynx (in ArtiSynth), epiglottal level**.**

Figure. 6. Vertical convex ultrasound**.**   **Figure. 7**. Paracoronal laryngeal section (in ArtiSynth).
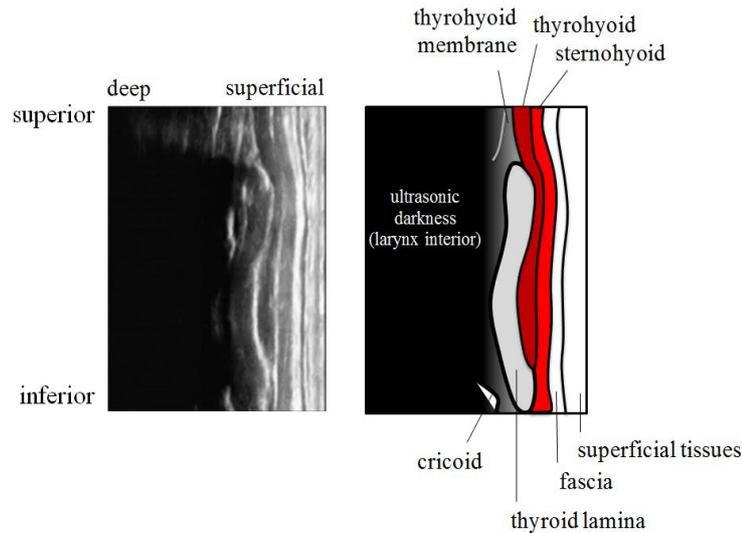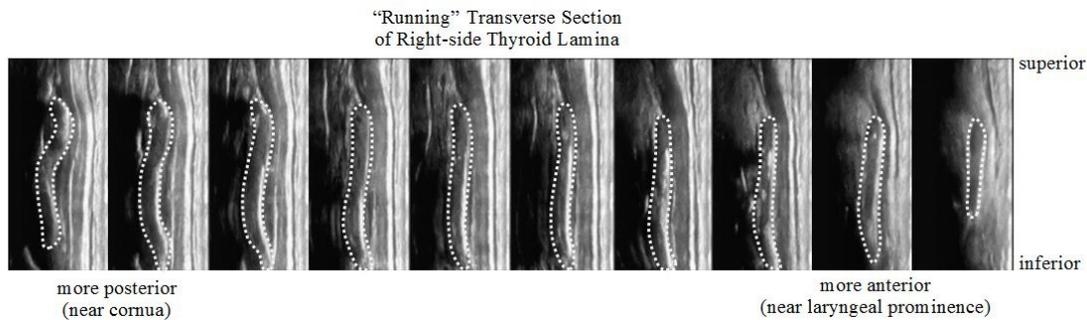
Figure. 8. Vertical flat ultrasound.

Figure. 9. Vertical flat ultrasound, running transverse section.

General purposes for the App

The purpose of this free App is to introduce the general public as well as specialized users of IPA symbolization, via iPhone/iPad technology, to the inventory of possible speech sounds of the languages of the world and to how each sound is physically articulated. Each sound category can be listened to and viewed in the form of close-up oral-endoscopic videos of the vocal tract. Pharyngeal/Epiglottal and Glottal (Laryngeal) articulations are also accompanied by laryngeal ultrasound images, which can be compared with the laryngoscopic videos of the same sound productions. Ultrasound images of this region of the vocal tract have not been included in any other database of articulatory categories that we know of. The ultrasound images for these lower-vocal-tract articulations may be compared, for example, with the lingual ultrasound images for the sounds of the phonetic chart contained in the Carnegie Trust Ultrasound Tongue Imaging Resource (http://www.seeingspeech.arts.gla.ac.uk/uti/). (Glasgow, Queen Margaret, Strathclyde, Edinburgh, and Aberdeen Universities) The notion of how users can be led to interpret and compare the ultrasound images in the database will be explored and discussed.

References

Ball, Martin J. 1991. Computer coding of the IPA: Extensions to the IPA. *Journal of the International Phonetic Association* 21, 36-41.

Duckworth, Martin, George Allen, William J. Hardcastle & Martin J. Ball. 1990. Extensions to the International Phonetic Alphabet for the transcription of atypical speech. *Clinical Linguistics and Phonetics* 4, 273-280.

Esling, John H. 1996. Pharyngeal consonants and the aryepiglottic sphincter. *Journal of the International Phonetic Association* 26, 65–88.

Esling, John H. 1999. The IPA categories 'Pharyngeal' and 'Epiglottal': Laryngoscopic observations of pharyngeal articulations and larynx height. *Language and Speech* 42(4), 349–372.

Esling, John H. 2005. There are no back vowels: The laryngeal articulator model. *Canadian Journal of Linguistics* 50, 13–44.

Esling, John H. 2010. Phonetic notation. In William J. Hardcastle, John Laver & Fiona E. Gibbon (Eds.), *The Handbook of Phonetic Sciences,* 2nd ed. Oxford: Wiley-Blackwell, 678-702.

Moisik, Scott R. 2013. *The Epilarynx in Speech.* PhD dissertation, University of Victoria.

Moisik, Scott & Bryan Gick. 2013. The quantal larynx revisited. *Proceedings of Meetings on Acoustics* 19 (060163).

Poster 7

Ultrasound in language documentation: Some observations on lingual movement in Lopit vowel contrasts

Rosey Billington

This paper presents some preliminary ultrasound-based observations on the articulation of vowel contrasts in Lopit, an un(der)described Eastern Nilotic language traditionally spoken in South Sudan. The present investigations form part of a larger study which aims to document the phonology of Lopit and investigate the acoustics, articulation and perception of the vowel contrasts in the language.

Existing materials on Lopit note that the language has a contrast between two sets of vowels on the basis of the phonological feature 'Advanced Tongue Root', or ATR (Turner 2001). This feature is widely attested in the vowel inventories of Niger-Congo and Nilo-Saharan languages, and is held to distinguish vowels of a similar height, backness and rounding through advancement of the tongue root in contrast to non-advancement (or retraction), conventionally +ATR and –ATR. However, articulatory investigations of lingual movement have only taken place for a small number of languages with this type of contrast attested. While there is good evidence that movement of the tongue root (combined with pharyngeal expansion) does occur when speakers of Niger-Congo languages of West Africa produce these types of vowel contrasts (e.g. Hudu, Miller & Pulleyblank 2009), less is known about lingual movement for speakers of Nilo-Saharan languages of East Africa.

Some early observations (Jacobson 1978) have suggested that for Nilo-Saharan languages, the gesture that accompanies this type of vowel contrast may be less uniform than that found in West African Niger-Congo languages, and may entail movement of the tongue body rather than the tongue root.

Acoustic analyses of Lopit vowels to date have shown that there are clear differences in the acoustic characteristics each vowel set; the vowels that have been noted as +ATR tend to have lower F1 values compared to the –ATR vowels, and there are some additional differences in duration and F2 values for particular vowel pairs. However, given that different lingual gestures can give rise to similar types of acoustic effects, instrumental investigation is required to see what sort of lingual movement speakers use to achieve these effects. Some preliminary results are presented regarding the relationship between the acoustics of Lopit vowels and the articulation of these vowels as observed with ultrasound imaging, together with some discussion about using these sorts of techniques in the early stages of language documentation.

References:

Hudu, Fusheini, Amanda Miller & Douglas Pulleyblank. (2009). Ultrasound imaging and theories of tongue root phenomena in African languages. In Peter K. Austin, Oliver Bond, Monik Charette, David Nathan & Peter Sells (eds.), Proceedings of Conference on Language Documentation and Linguistic Theory 2. London: SOAS.

Jacobson, Leon Carl. (1978). DhoLuo Vowel Harmony: A phonetic investigation. UCLA Working Papers in Phonetics, 43.

Turner, Darryl. (2001). Lopit phonology: an outline of the phonology of Lopit words with a proposal for a tentative working orthography. Unpublished manuscript, SIL-Sudan.

Poster 9:

Segment Constants and Speaker Constants in English Coronal Obstruents

Simon Gonzalez, Mark Harvey, Susan Lin, Katherine Demuth

Key words: articulation, coronals, tongue body, tongue blade, tongue tip, ultrasound

This paper examines the articulation of the coronal obstruents /t, s, θ, tʃ, ʃ/. Previous studies have found that there is variation in the production of coronals (Dart, 1991; Delattre & Freeman, 1968; Reichard et al., 2012; Stone et al., 2012). These studies examined single segments (/r/, /s/). We compared segments with different places and manners of articulation, with the aim of isolating the articulatory correlates of the various coronal places in English. We used ultrasound to examine tongue motion in real time (Bressmann, 2008; Epstein & Stone, 2005). We provide evidence that there is variation in the articulation of coronal obstruents, but this variation is not unconstrained. We found articulatory constants for each segment that held across all speakers. Further, we found that variation was almost entirely inter-speaker and only minimally intra-speaker.

We recruited ten female native speakers of Australian English, and recorded them in a sound-attenuated testing room. We asked them to listen to pre-recorded sentences and then to repeat each sentence five times. We placed each target word in the carrier sentence "Please, utter X publically". The target words were *tack*, *sack, Thack, Chack* and *shack*, and with the target sounds /t, s, θ, tʃ, ʃ/ in onset position, respectively. All target sounds were therefore flanked by the vowels, /ə/ and /æ/. We recorded ultrasound mid-sagittal plane images (29.97fps) of target segments with a portable Sonosite 180 Plus ultrasound machine with a C11/7-4 MHz 11-mm broadband curved array transducer. We then fitted tongue contours by using EdgeTrak. We stabilized the probe with an Alan's headset. This headset allows the probe to be fixed and capture the ultrasound from a constant position in all utterances.
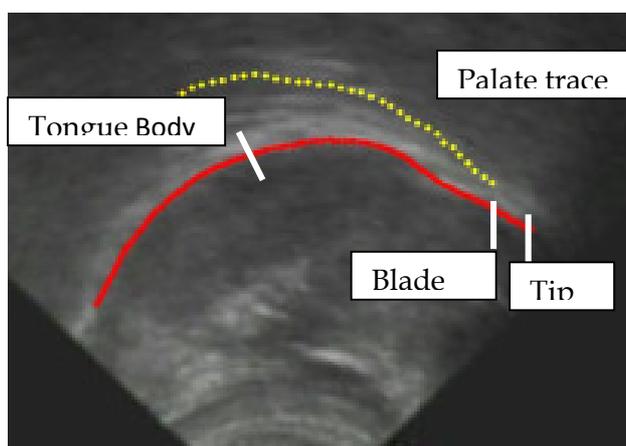

Figure 1. Sections of the tongue for the analysis of coronal sounds

We observed significant differences in the patterns of movement between segments and between speakers in terms of three areas of the ultrasound images. These three areas were (1) the most anterior portion of the images (approx. 5mm), (2) the next most anterior portion of the image (approx. 15mm) and (3) the remainder of the image. We refer to Area 1 as the tip, Area 2 as the blade and Area 3 as the body. We use these terms as the patterns of movement that we observed in these three divisions corresponded to the movements traditionally ascribed to the tip, blade, and body for coronal obstruents.

For each sound, we located the frame which corresponded most closely to maximum constriction (MC). In the MC frame, we examined the position of the tongue body in relation to the front sections: the blade and the tip. We then tracked three frames (100ms) prior to MC and examined which portion of the tongue moved to form MC.

For each place of articulation, there was a segment-specific consistent pattern across speakers of movement into MC. In the case of /t/ and /s/, the tip moved upwards and forward to attain MC. In the case of /θ/, the blade moved forward to attain MC. In the case of /tʃ/ and /ʃ/, the blade moved upwards to attain MC. Then depending on the place of articulation, there was inter-speaker variation in the positioning of the tongue body at MC. For /tʃ/ and /ʃ/ there was no variation – the tongue body was always raised. For /t, s, θ/ there was inter-speaker variation in whether the tongue body was raised or not. For seven speakers the body was raised (see Figure 2). For two speakers the body was lowered in relation to the front part of the tongue (see Figure 2). Only one speaker showed a mixed pattern, where the tongue body was consistently raised for /t/, but not for /s/ and /θ/.

The critical factor predicting inter-speaker consistency appears to be whether the characteristic movement into MC requires also raising the tongue body or not. Raising the blade requires raising the body, and so /tʃ/ and /ʃ/, which both require tongue blade raising, show no inter-speaker variation in this regard. The characteristic movements toward MC for the other coronal obstruents do not require raising the blade. Consequently, there is no segmental conditioning on the position of the tongue body. Our data shows that tongue body position for these coronals does not vary randomly, but rather is constant for a given segment, and usually constant for a given speaker.
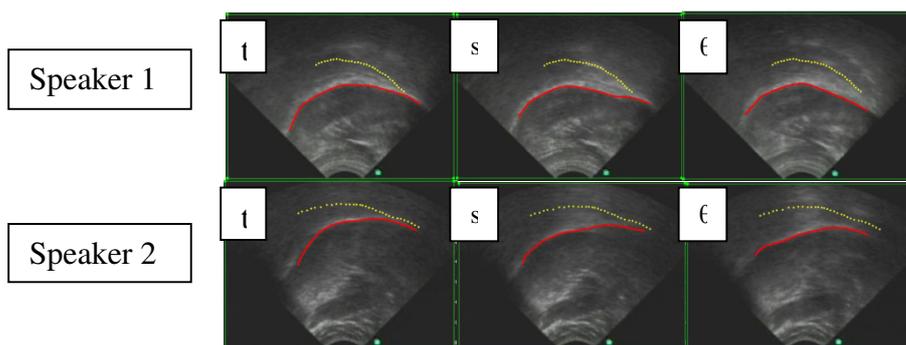

Figure 2. MC for / t, s, θ/ produced by speaker 1 showing raised tongue body and speaker 2 showing depressed tongue body.

References

Bressmann, T. (2008). *Quantitative Assessment of Tongue Shape and Movement Using Ultrasound Imaging.* Paper presented at the Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology, Somerville, MA.

Dart, S. (1991). Articulatory and Acoustic Properties of Apical and Laminal Articulations. *UCLA Working Papers in Phonetics, 79.*

Delattre, P., & Freeman, D. (1968). A dialect study of American R's by X-ray motion picture. *Linguistics, 44*, 29-68.

Epstein, M. A., & Stone, M. (2005). The tongue stops here: Ultrasound imaging of the palate. *Journal of the Acoustical Society of America, 118*(4), 2128-2131.

Reichard, R., et al. (2012). *Motion of apical and laminal /s/ in normal and post-glossectomy speakers.* Paper presented at the Proceedings of the Acoustics 2012 Hong Kong Conference and Exhibition, Hong Kong, China.

Stone, M., et al. (2012). Frequency of Apical and Laminal /s/ in Normal and Post-glossectomy Patients. *Journal of Medical Speech Language Pathology, 20(4),* 106-111

Probe Skewing Effects & Kinect Head-tracking in Tongue Contour Imaging

Samuel Johnston, Daniel Brenner, Rolando Coto, and Diana Archangeli

The University of Arizona, Tucson, Arizona, USA

The position of the ultrasound probe in relation to the position of the head of the subject (the tilt of the head or probe) may affect the quality and comparability of ultrasound tongue images, compromising the interpretability of the data. The challenge of maintaining alignment of the probe with a subject's head is a substantial one, often involving immobilization of the head or fixation of the probe beneath the subject's chin, using helmets, cages, and other mechanical systems of varying complexity (see Stone and Davis (1995) on HATS; Whalen et al. (2005) on OptoTrak and HOCUS; and McLeod and Wrench (2008) on helmets). In the field especially (but also in the laboratory), many subjects, particularly the very young and the very old, may be disinclined to have themselves manipulated in these ways, and sometimes find the physical apparatuses imposing, uncomfortable, too heavy, or even frightening. At the same time, the effects of probe skewing that such contraptions are meant to prevent have yet to be assessed quantitatively.

This study compares control images (no tilt) and images gathered with a tilted head to determine the extent to which control images differ from tilt images. We measure the degree of tilt via Microsoft Kinect, using synchronization software developed in our lab. Kinect provides a low-cost non-invasive alternative to systems like OptoTrak (Whalen et al., 2005) and is a portable system with precise measurement of movements (unlike the image-mixing system used with Palatoglossatron (Baker, 2005)).

Subjects in the experiment produce each of the sounds /æ, ɑ, ɹ, l, t, k, s, ʃ, θ /, selected to induce maximally diverse English language tongue gestures. These extremes of articulation cast a wide net for capturing effects of probe skewing.

Subjects produce and hold each of the target sounds through an arc of head roll, from one side, through 0°roll, to the other side, and back to 0°roll. We will report a fitted distance function of tilted images to 0°images with confidence intervals.

References

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

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(1):23 – 29.

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.

Whalen, D. H., Iskarous, K., Tiede, M., Ostry, D., Lehnert-LeHouillier, H., and Vatikiotis-Bateson, E. 2005. The Haskins optically corrected ultrasound system (HOCUS). *Journal of Speech, Language, and Hearing Research*, 48(3):543.

A Comparison of Methods for Tongue Shape Analysis

Katherine M. Dawson1, Douglas H. Whalen1, 2, Mark K. Tiede1, 2

1City University of New York, 2Haskins Laboratories

There are many possible methods for analyzing tongue shape data obtained from ultrasound, with the choice usually driven by the research questions of the study. The aim of the present study is to

meaningfully describe through the use of a quantification parameter the differences in complexity between tongue shapes from utterances of various types, and to test whether a useful range for such a complexity parameter can be captured by differing candidate methods for a normal population. The most promising methods can then be used in further work to index differences in tongue shape complexity between normal and clinical populations.

Different analysis techniques will be applied to a restricted set of data, recorded from Québécois French speakers. Two of these techniques are based on Procrustes analysis (Goodall, 1991), and relate to deformation and rotation of the shape, and two are based on polynomial fitting (Stolar & Gick, 2013; Van Otterloo, 1988; Young, Walker, & Bowie, 1974).

Results will be described in terms of how well the methods differentiate more complex tongue shapes (i.e. those with multiple inflections / extreme constrictions) from simpler ones. The analysis methods will also be compared with respect to their resilience to the limitations of ultrasound imaging, such as minor movements of the probe and data variability. The results of this study may also indicate modifications that can be made to existing methods for improved application to the analysis of tongue shape complexity.

References

Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society, 53(*2), 285-339.

Stolar, S. & Gick, B. (2013). An index for quantifying tongue curvature. *Canadian Acoustics, 41*(1), 11-15.

Van Otterloo, P.J. (1988). *A contour-oriented approach to digital shape analysis* (Doctoral Dissertation). Delft University of Technology, Delft, The Netherlands.

Young, I. T., Walker, J. E., & Bowie, J. E. (1974). An analysis technique for biological shape. I. *Information and control, 25*(4), 357-370.

Poster 15:

Quantitative analysis of ultrasound data collected with and without head stabilisation

Natalia Zharkova and Nigel Hewlett

Clinical Audiology, Speech and Language Research Centre, Queen Margaret University

Ultrasound research on speech production in typical speakers has long used quantitative measurements (Stone 2005). By contrast, previous studies using ultrasound tongue imaging with clinical populations have generally reported qualitative information on tongue movement (e.g., Bacsfalvi & Bernhardt 2011; Bressmann et al. 2011). The development of quantitative measures for use in the clinic has largely been hindered by the need to stabilise the ultrasound transducer in relation to the head of the speaker. The aim of the present study is to explore the applicability of quantitative measurements of stabilisation-free tongue movement data, by comparing ultrasound data collected with and without head stabilisation. The stabilisation device used in the study is a headset holding the transducer (Articulate Instruments Ltd 2008).

Ultrasound tongue movement data synchronised with acoustic data from ten adolescent speakers (aged between 13 years 0 months and 13 years 11 months) were recorded and analysed. The speakers produced the syllables /pa/ and /pi/ in the carrier phrase "It's a ..., Pam". Each target was repeated five times, and the tokens were presented in random order. The ultrasound fame rate was 100 Hz. Recordings were made in a sound-treated studio. Each participant performed the task two times. The first time was recorded while the participant wore the headset, and the second time no headset was worn, instead the transducer was hand-held by the experimenter (see Figure 1). The data collection without the headset was video recorded.
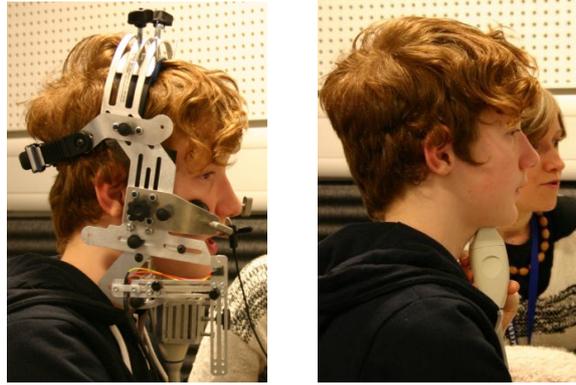
Figure 1. Illustrations of recordings: participant wearing a headset (on the left); transducer hand-held by experimenter (on the right).

Midsagittal tongue curves at /p/ mid-closure were compared across the two conditions (headset versus non-headset) and across vowel contexts (/a/ versus /i/). First, the Nearest Neighbour calculations (Zharkova & Hewlett 2009) were performed. They involved computing across-set (AS) distances between two sets of curves for /p/ (one set in the context of /a/ and one set in the context of /i/) and within-set (WS) distances for each of the two sets of curves. We hypothesised that there would be an effect of condition on both AS and WS distances. We expected to observe a significant vowel-related difference between the two sets of consonant tongue curves (i.e., significantly greater AS distances than WS distances for each vowel context) in the headset condition, and no such difference in the no-headset condition. Next, two measures were taken which characterise the tongue shape through ratios based on a single curve (Zharkova 2013): Dorsum Excursion Index (DEI) and Tongue Constraint Position Index (TCPI). We predicted no effect of condition and a significant effect of vowel context on both indices.

There was a significant effect of condition on AS and WS distances ($p < 0.01$). WS distances were consistently larger in the no-headset condition for every speaker, while for AS distances there was a noticeable inter-speaker variation. The effect of condition on DEI was not significant for the data pooled across vowel contexts and for the data from the context of /i/ only, but the effect was significant for the data from the context of /a/ (at $p < 0.01$). The effect of condition on TCPI was not significant either for the data pooled across the two vowel contexts or for the data from the two vowel contexts analysed separately. For both headset and no-headset conditions, a significant vowel-related difference in the consonant tongue curves was observed when comparing AS distances to WS distances ($p < 0.01$). The vowel context effect was significant in both headset and no-headset conditions for both DEI and TCPI ($p < 0.01$), with all speakers showing larger values in the context of /i/ for both indices.

As predicted, the measures based on calculating absolute distances between tongue curves were affected by whether or not the speakers wore the headset. Measures based on single curves have previously been found to be unaffected by changes in transducer placement (Ménard et al. 2012), and our findings were largely consistent with this pattern. However there was a vowel-specific difference for DEI, with the no-headset condition in the context of /a/ showing, on average, smaller values than the headset condition. This difference was not uniform across speakers. Only five speakers had smaller DEI values in the no-headset condition in this vowel context, however the difference across conditions tended to be particularly large in four of these speakers. Future work will include analysing tongue curves for other consonants in the headset and no-headset conditions, to investigate whether the same effect on DEI in the context of /a/ is observed regardless of the consonant identity.

TCPI was the only measure unaffected by the headset/no-headset condition in the two vowel contexts pooled and in each vowel context analysed separately. This finding suggests that TCPI can be used for comparing tongue behaviour in speakers who can wear the headset and speakers who cannot wear it, for example, young children. Our results also show that relative measurements, such as comparisons of index values within speaker across vowel contexts, provide more consistent results across speakers than comparisons of absolute index values across speakers. Variability in absolute values found in this

study is in agreement with previous articulatory findings (e.g., Gibbon et al. 2007; Liker & Gibbon 2008).

References

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

Bacsfalvi, P. & Bernhardt, B. M. (2011). Long-term outcomes of speech therapy for seven adolescents with visual feedback technologies: ultrasound and electropalatography. *Clinical Linguistics & Phonetics*, **25**, 1034-1043.

Bressmann, T., Radovanovic, B., Kulkarni, G.V., Klaiman, P. & Fisher, D. (2011). An ultrasonographic investigation of cleft-type compensatory articulations of voiceless velar stops. *Clinical Linguistics & Phonetics*, **25**, 1028-1033.

Gibbon, F., Lee, A. & Yuen, I. (2007a). Tongue-palate contact during bilabials in normal speech. *Cleft Palate-Craniofacial Journal*, **44**, 87–91.

Liker, M. & Gibbon, F. (2008). Tongue palate contact patterns of velar stops in normal adult English speakers. *Clinical Linguistics and Phonetics*, **22**, 137–148.

Ménard, L., Aubin, J., Thibeault, M., & Richard, G. (2012). Comparing tongue shapes and positions with ultrasound imaging: a validation experiment using an articulatory model. Folia Phoniatrica et Logopaedica, 64, 64–72.

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

Zharkova, N. (2013). Using ultrasound to quantify tongue shape and movement characteristics. *The Cleft Palate-Craniofacial Journal*, **50**, 76-81.

Zharkova, N., & Hewlett, N. (2009). Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and /ɑ/. *Journal of Phonetics,* **37**, 248-256.

Poster 17:

Investigation of gestural coordination in the production of consonant clusters in children with typical development and deviant language

Aline Mara De Oliveira Vassoler[1], Larissa Cristina Berti[2]

1Department of Linguistics – Universidade Estadual Paulista (Unesp – Rio Preto)
2 Department of Linguistics (Oralidade e Letramento) -Universidade Estadual Paulista (Unesp – Rio Preto e Marília)

Introduction

In light of Phonology Gestual the production of canonical syllables, like CV syllables, involves a pattern of coordination more stable between articulatory gestures incorporation of the consonant and vowel syllables. While the production of syllables more complexes, such as CCV, involves a gestural coordination pattern more complex and less stable. Although there are some reports in the literature that children often tend to simplify/reduce the syllabic pattern for CCV to CV, studies using acoustic-phonetic analysis have identified the presence of covert phonic contrasts in the speech of children who have not yet acquired effectively CCV syllable, which indicates that there is an effective reduction of CCV standard for the standard CV ((NAM; SALTZMAN, 2003; GOLDSTEIN et al., 2007). The objective of the study is to investigate and describe, the perspective of dynamic (FonGest), the pattern of gestural coordination imbricated in the production of syllabic type CCV versus CV of children with typical and atypical phonological development.

Methodology

And even if the articulatory analysis provides these children the rescue of coordinations gestural

imbricated in the production of consonant clusters. These project will analyse five children with phonological disorder (PD) and difficulties in the production of syllabic pattern
CCV. Five children with typical language development will be analysed, according to their ages, gender and education level. The ultrasound production of each children will be analysed. Both ultrasound images of movements of the tongue and the audio signals are simultaneously obtained by means of the portable ultrasound device, coupled with appropriate equipment and software analyzed by AAA (Advanced Articulate Assistant) (ARTICULATE INSTRUMENTS, 2012).

Analysis

The results will be analyzed qualitatively as the gestural descriptors proposed by FonGest. In addition to contributions to the study of phonics acquisition it is also expected the discussion about the primitive phonological.

References

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

GOLDSTEIN, L; CHITORAN, I.; SELKIRK, E. Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber. ICPhS 16, p. 2153-2156, 2007.

NAM, H.; SALTZMAN, E. A Competitive, Coupled Oscillator Model of Syllable Structure. Paper presented at the Proceedings of the XIIth International Congress of Phonetic Sciences, Barcelona, Spain, 2003

Poster 19:

Tongue contour in /ʃ/ sound for Brazilian Portuguese-speakers with SSD: a pilot study using ultrasound

Haydée Fiszbein Wertzner, Danira Tavares Francisco, Luciana de Oliveira Pagan-Neves

Department of Physiotherapy, Communication Sciences and Disorders and Occupational Therapy from the Medical School. University of São Paulo, Brazil.

Background:

The specific use of ultrasound images as a complementary analysis to the diagnosis of speech sound disorders (SSD) is a recent issue in the literature. **Aim**: Describe the tongue shape for /s/ and /ʃ/ in three different groups of children with and without SSD.

Methods

The six participants were divided into three groups: Group 1- two typically developing children, Group 2- two children with SSD presenting any other phonological processes but not the ones involving the production of the /ʃ/ and Group 3- two children with SSD presenting any phonological processes associated to the presence of the phonological process of palatal fronting (these two children produced /ʃ/ as /s/) aged between 5 and 8 years-old, all speakers of Brazilian Portuguese. Data were the words /'ʃavi/ (key) and /'sapu/ (frog). Ultrasound imaging of the tongue shape during the productions was analyzed.

Results and Conclusion

The analysis of the tongue contour indicated evidences that both /s/ and /ʃ/ were produced using distinct tongue contours which permitted the identification of the sounds. The tongue contour for /s/ and /ʃ/ in children with SSD presenting the phonological process of palatal fronting was similar demonstrating that their production was undifferentiated. Therefore, we can conclude that the use of

the ultrasound applied to the speech analysis was effective to confirm the perceptual analysis of the sound made by the speech pathologist.

Poster 21:

An Exploratory Ultrasound Investigation of Emphatic Articulation in Cairene Arabic

Natalia Lapinskaya

University of Toronto

Secondary articulation of 'emphatic' consonants of Arabic has been described as lowering of the tongue in the palatal region and retraction of the tongue root towards the pharynx (pharyngealization; Ali & Daniloff, 1972). Previous investigations of emphatic articulations (including ultrasound, such as Zeroual et al., 2011) have been primarily qualitative in nature. This study presents an exploratory quantitative investigation of emphatic articulation, employing Smoothing Spline Analysis of Variance (SS-ANOVA; Davidson, 2006) to statistically evaluate the differences in position between individual sections of the tongue during the articulation of plain and emphatic consonants.

A female native speaker of Cairene Arabic was recorded pronouncing multiple repetitions of meaningful words containing plain /t, d, s, k/ and emphatic /tˤ, dˤ, sˤ, kˤ/ in the context /a_a/. 86 tokens were analyzed in total, 10-12 per consonant. Scans were performed with a 7.5 MHz *SeeMore* USB ultrasound probe by *Interson*, with scanning depth set to 10cm and a scanning rate of 15 frames per second. The audio signal was captured using an AT831b lavalier microphone and a *Sound Devices* USBPre2 pre-amplifier with a sampling rate of 48 kHz. Both were connected via USB to an *Acer Aspire* 5755 laptop computer with *SeeMore* software (version 1.03.02) installed. Tongue contours from frames corresponding to the greatest constriction during the consonant interval were traced semi-automatically using *EdgeTrak* software (Li et al., 2005), manually adjusted when necessary, and exported for analysis as a set of 100 XY coordinates. SS-ANOVA was used to determine which sections of the tongue contours differ significantly by comparing 95% Bayesian confidence intervals constructed around smoothing splines best fitting the plain and emphatic data sets for each consonant (Davidson, 2006).
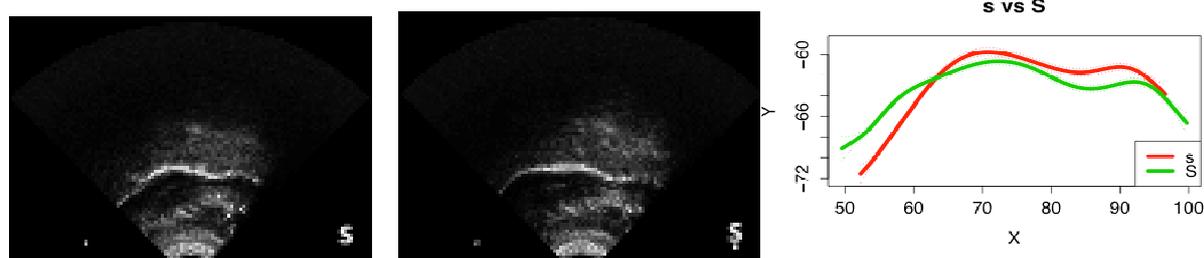

Figure 1. Sample ultrasound frames & SS ANOVA comparison for /s/ (left) and /sˤ/ (='S'). Tongue tip is on the right.

The results showed significant lowering of the tongue in the palatal region during the production of all emphatics, compared to non-emphatic counterparts. The posterior tongue dorsum was significantly retracted, as illustrated in Figure 1. This is consistent with previous research involving ultrasound (Zeroual et al., 2011), cinefluorography (Ali & Daniloff, 1972), videofluoroscopy (Al-Tamimi et al., 2009), and other methods. The position of the most anterior section of the tongue was not found to be significantly different for any plain/emphatic pair. The shape of the tongue, aside from the retracted portion, was found to be nearly identical for plain and emphatic coronal segments (/t, d, s/). This is in contrast with the view that the point of tongue-palate contact is retracted for the emphatics, as

previously suggested by Al-Ani (1970), among others. In sum, the results of this study suggest that quantitative ultrasound analysis of Arabic emphatics can be used successfully to evaluate previous generalizations about these sounds.

References

Al-Ani, S. H. (1970). *Arabic phonology: An acoustical and physiological investigation*. The Hague: Mouton.

Al-Tamimi F., Alzoubi F., Tarawnah R, A. (2009). Videofluoroscopic Study of the Emphatic Consonants in Jordanian Arabic. *Folia Phoniatrica et Logopaedica*, 61, 247-253

Ali, L. & Daniloff, R. (1972). A contrastive cineflurographic investigation of the articulation of emphatic-non emphatic cognate consonants. *Studia Linguistica,* 26, 81-105.

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.

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

Zeroual, C., Esling, J. H., & Hoole, P. (2011). EMA, endoscopic, ultrasound, and acoustic study of two secondary articulations in Moroccan Arabic. In Hassan, Z.M., Heselwood, B., (Eds.) *Instrumental Studies in Arabic Phonetics* (pp. 277-301). Philadelphia: John Benjamins North America.

Poster 23:

Glottal Stop in Arabic: The Insufficiency of a Landmark Underspecification Account

Laura Ryals

Georgetown University

The glottal stop is phonetically enigmatic. In many languages, the articulatory singularity of glottal stop is accompanied by behavioral and distributional singularity. In Marianne Borroff's 2007 dissertation, "A Landmark Underspecification Account of the Patterning of Glottal Stop", she develops a proposed causal relationship between the two. She argues that glottal stop features fewer clues ("landmarks") in its acoustic signal that can inform the listener as to what gesture is being produced. This absence of certain landmarks prevents glottal stop from being able to properly align with adjacent segments, resulting (she argues) in the observed behavior and distribution.

Yet, glottal stop in some languages, including Arabic, patterns as though it *does* in fact possess the landmarks that Borroff claims it does not. Her proposed solution is that glottal stop in these languages is accompanied by a secondary articulation, further upstream. This secondary articulation is the component possessing the needed landmarks and allowing gestural coordination to occur.

However, by utilizing nasoendoscopic, videofluoroscopic, and acoustic data collection methods, Heselwood et al. (2011) have demonstrated the lack of such an articulation during the pronunciation of glottal stop in Jordanian Arabic. Shahin (2011) has likewise used nasoendoscopy to demonstrate a similar lack in Palestinian Arabic.

In this study, I aimed to widen the scope of such research by looking for the presence or absence of a secondary articulation during the pronunciation of glottal stop in Syrian Arabic. I used pharyngeal ultrasound to image the position of the tongue root during the pronunciation of glottal stop, during the pronunciation of the voiced pharyngeal approximant, and during the pronunciation of the alveolar consonants [s], [t], and [d], along with their pharyngealized counterparts.

By overlaying a trace of the tongue root position during the articulation of a pharyngealized alveolar consonant with that of its non-pharyngealized counterpart, I was able to see what a typical secondary articulation looked like for this speaker. Such an overlay then provided a comparative basis for the secondary articulation proposed during glottal stop. For example, Figure 1 is a conflation of the traces of the tongue root position during the articulation of [t] (line) and [tˤ] (dash); Figure 2 is a conflation of

the traces of the tongue root during the articulation of [ʔ] (line) and [t] (dash); and Figure 3 is a conflation of the traces of the tongue root

during the articulation of [ʔ] (line) and [tˤ] (dash).

| Figure 1 | Figure 2 | Figure 3 |
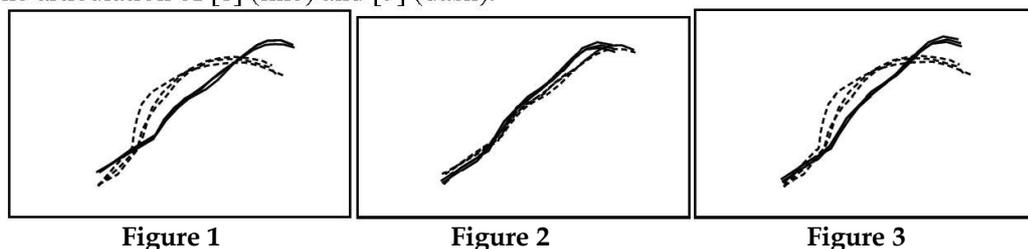
Together, these figures suggest that glottal stop is as equally lacking in a secondary articulation as is the non-pharyngealized alveolar consonant. While these current results are primarily qualitative, my next step is to apply statistical methods of curve comparison analysis to these images to add a quantitative component.

Establishing a lack of secondary articulation during glottal stop in Arabic would demonstrate that it is indeed an exception to Borroff's account of the patterning of glottal stop cross-linguistically, which would provide further insight into the unique behavior of this consonant.

Poster 25:

Describing the Pharyngeal Constriction Gesture in Turkish Rhotics

Sarah Hamilton, Suzanne Boyce

University of Cincinnati

In recent years, the increasing availability of ultrasound technology has allowed clinicians to use ultrasound for speech therapy, increasing the subject pool for research on the articulatory characteristics of the /r/ sound. Such ultrasound speech therapy studies have focused mainly on the /r/ sound in English, reported to be the both most difficult sound for children to produce (St. Louis, Ruscello, & Lundeen, 1992) and the most difficult sound for clinicians to remediate (Shuster, Ruscello, & Toth, 1995). Articulatory imaging studies have revealed why English /r/ may be so difficult to produce: The /r/ sound requires differentiated movements of the front and back parts of the tongue. While the front of the tongue may be characterized broadly into "bunched" and "retroflexed" shapes, there are many distinct shapes that it can take during an acceptable /r/ production in English. However, together with the primary constriction from the tongue body, the root must move in a gesture of pharyngeal constriction that assists in lowering the third formant that is characteristic of the perceptually-correct/r/ sound (Lindau, 1985; Espy-Wilson, Boyce, Tiede, Holland, & Choe, 2008). In a descriptive study of nine clients enrolled in ultrasound therapy for the /r/ sound, clients' error productions had inconsistent pharyngeal constriction, but pharyngeal constriction was present in all their correct productions (Hamilton, Boyce, Rivera Campos, McNeill, & Schmidlin, 2013). These results provide more evidence that movement in the pharyngeal space is a secondary articulatory characteristic of the /r/ sound in English.

Rhotic sounds present challenges to speech pathologists who work with different languages, too. Children who have speech and language disorders commonly misarticulate these sounds, even in languages with different phonological rule systems, such as in Turkish (Topbaş, 2006). Turkish

provides a testing ground for our observation that pharyngeal constriction is part of the acoustic signature of the low F3 in the /r/ sound, by moving away from a language-specific phonology and attempting to define a general rhotic-class characteristic. In addition, Turkish is a language known to have vowel and consonant-harmony, in which underlying phonemes have different phonetic (and thereby articulatory) realizations. Turkish, therefore, may serve to test the limits of variation in rhotic articulation, as to what articulatory parameters can be changed and yet still remain a rhotic sound. However, most studies of Turkish rhotic sounds have focused on their acoustical characteristics and not on using ultrasound to image the articulatory characteristics of these sounds.

In this study, two Turkish speakers were recorded saying a list of Turkish words, using an ultrasound machine with a simultaneous video and audio feed. Acoustic and articulatory observations indicated that pharyngeal constriction was present during certain rhotic sounds (such as the voiceless fricated flap that occurs in word-final position, /ř/), but was variably present during the intervocalic allophone flap (/ɾ/). In addition, the acoustic profile of the fricative rhotic /ř/ showed the characteristic low third formant of the English /r/. These results strengthen the case for the lowered third formant resulting from a combination of primary constrictions in the oral cavity and secondary constriction in the pharyngeal cavity.

References

Espy-Wilson, C. & Boyce, S., Jackson, M., Narayanan, S., & Alwan, A., (2000) "Acoustic Modeling of American English /r/", *Journal of the Acoustical Society of America*, 108,343-356.

Hamilton, S., Boyce, S., Rivera Campos, A., McNeill, B., & Schmidlin, S. (2013). Articulatory and acoustic evidence for the role of pharyngeal constriction in the /r/ sound. Poster presented at Ohio Speech, Language, and Hearing Association conference, Columbus, Ohio.

Lindau, M. (1985). The story of /r/. In Phonetic Linguistics. Ed. Fromkin. Orlando, Academic Press: 57-168.

Shuster, L. I., Ruscello, D. M., & Toth, A. R. (1995). The use of visual feedback to elicit correct/r. *American Journal of Speech-Language Pathology*, 4(2), 37.

St. Louis, K. O., Ruscello, D. M., & Lundeen, C. (1992). Coexistence of communication disorders in schoolchildren. *Pathology, 1,* 29-34.

Topbaş, S. (2006). Does the speech of Turkish speaking phonologically disordered children differ from that of children speaking other languages? *Clinical Linguistics & Phonetics*, *20*(7-8), 509–522. doi:10.1080/02699200500266331

Poster 27:

Contextual patterns of /l/-darkening in accents of English

Danielle Turton

University of Manchester

The phenomenon of /l/-darkening, whereby /l/ is produced with a delayed tongue-tip gesture, exhibits a remarkable amount of variation depending on its position in the word or phrase. Traditionally, it is said that light [l] occurs in onsets (e.g. *light*) and dark [ɫ] in codas (e.g. *dull*; Giegerich 1992; Halle & Mohanan 1985), but many studies report varying distributions of light and dark /l/ in different morphosyntactic environments. Table 1 summarises the outcome of previous studies on /l/-darkening in American varieties, alongside the pattern reported in RP.

	*ligh*t	*yello*w	*heal-in*g	*hea*l *i*t	*hea*l	
RP	[l]	[l]	[l]	[l]	[ɫ]	Cruttenden (2008)
Am. Eng. 1	[l]	[l]	[l]	[ɫ]	[ɫ]	Sproat & Fujimura (1993)
Am. Eng. 2	[l]	[l]	[ɫ]	[ɫ]	[ɫ]	Olive et al. (1993)
Am. Eng. 3	[l]	[ɫ]	[ɫ]	[ɫ]	[ɫ]	Hayes (2000)

Table 1: /l/-darkening in different environments. Adapted from Bermúdez-Otero (2007)

Beyond this variation in morphosyntactic conditioning, some dialects have been reported to show no allophonic distinction between light and dark /l/. The variety spoken in Manchester is said to exhibit dark [ɫ] in all contexts (Cruttenden 2008; Kelly & Local 1986), whilst North-East /l/s are reported as being clear in all positions (Cruttenden 2008; Wells 1982). Although such reports are widespread in the existing literature, they are yet to be vindicated by instrumental articulatory evidence, or even acoustic data.

This paper presents ultrasound data collected to test whether the aforementioned dialects of British English indeed display the reported patterns of /l/-darkening, and whether British English dialects show any evidence of morphosyntactic conditioning of /l/-darkening, as previously reported for American English.

Speakers of RP, and speakers from Manchester and Middlesbrough were recorded producing /l/ in five contexts: word-initial, word-medial before a vowel in the same stem, word-medial before a suffixal vowel, word-final prevocalic, and phrase-final, corresponding to the headings in Table 1. The RP speaker illustrated in Figure 1 shows the pattern of /l/-darkening reported by Cruttenden (2008), with [ɫ] only in non-prevocalic position: the backed tongue body, reduced tongue-tip gesture, and retracted tongue root typical of [ɫ] are found in prepausal *hea*l only, and not in *hea*l *it*. This demonstrates that there exist varieties of RP that do display very limited /l/-darkening, with a phrase-level alternation between light [l] prevocalically, and dark [ɫ] phrase-finally.
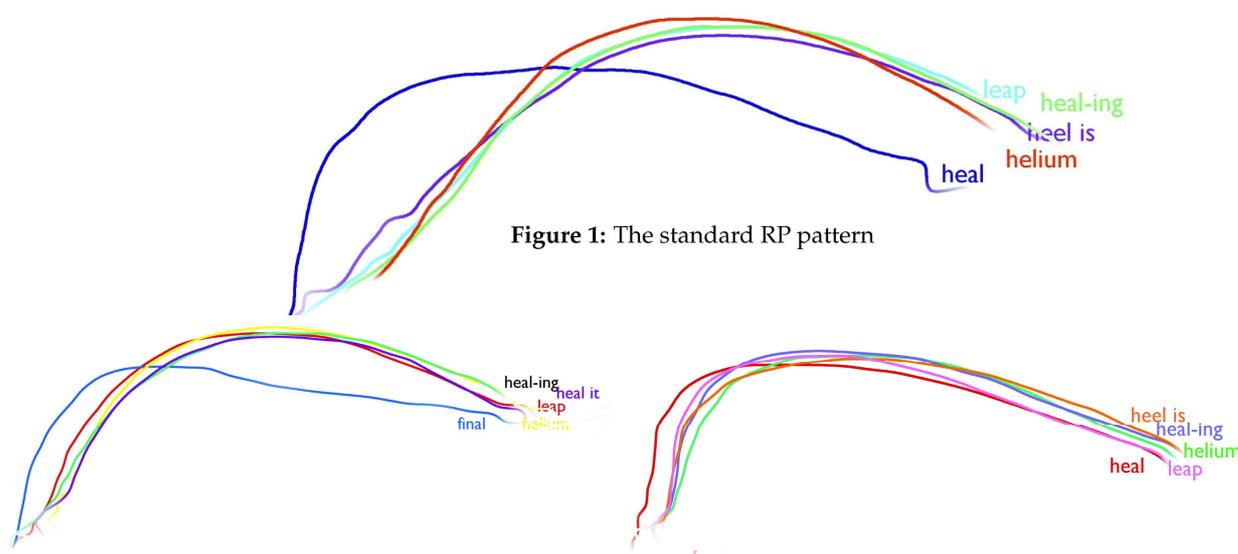

Figure 1: The standard RP pattern

Figure 2: The Middlesbrough pattern

Figure 3: The Mancunian pattern

The Middlesbrough speaker in Figure 2 shows the same distribution as RP, with phrase-final /l/ significantly backed in comparison with the four other phonological contexts (all statistical tests performed in the Articulate Assistant Advanced spline workspace with two-way t-test function; Wrench 2007). The Mancunian data (Figure 3) may appear to corroborate the claims of no allophonic distinction between initial and final /l/, however, the ultrasound imaging shows that phrase-final /l/

in the Mancunian data has marginal but significant tongue root backing (p < 0.05) compared with the other contexts.

The data from the three dialects provide hitherto absent instrumental evidence for a morphosyntactically conservative distribution of /l/-darkening, documenting an earlier stage of a well-reported sound change. Moreover, the fact that the three varieties show similarities in terms of **distribution**, but wide-ranging differences in terms of **realisatio**n raises important questions about the abstract nature of allophonic categories.

Poster 29:

An ultrasound study of Mandarin fricatives and affricate consonants

Xiu Ming & Didier Demolin

Gipsa-Lab, Université de Grenoble, Alpes

This presentation will describe the articulatory contrasts between alveolar, post-alveolar (retroflex), palatalized post-alveolar (alveo-palatal) fricative and affricate consonants in Mandarin using the utrasound technique. The aim is to compare tongue contours between the three sets of consonants (alveolar, post-alveolar, palatal and alveo-palatal) and to describe how they are articulated. The present state of the study study is based on recordings made with 3 Mandarin speakers using the following set of data (numbers indicate tones):

[tʂ] vs [ts] : [tʂang4] (账 account) vs [tsang4](藏 Tibet)
[tʂ] vs [ts] : [tʂang1] (张 family name) vs [tsang1](脏 dirty)
[tʂʻ] vs [ts] : [tʂʻeng2] (成 to become) vs [tseng2](层 floor)
[ʂʻ] vs [ts] : [ʂʻi1] (师 teacher) vs [si1](思 to think)

Ladefoged and Maddieson (1996) mention that sounds like '…[ʂ] does not involve the tip of the tongue being curled up and backwards into the palatal region as in the Dravidian subapical retroflex stops, nor does it have the apical post-alveolar shape that occurs in the Hindi retroflex stops'. For Ladefoged and Maddieson this sound is produced with the upper surface of the tip of the tongue, making it a laminal rather than an apical post-alveolar. The data recorded in the present study seem to confirm these observations for fricatives. However, they also suggest that the strategies of Mandarin speakers to produce sounds like [tz̢] are a bit more subtle. Indeed it seems that in this case the tongue is curled up but without involving a backward movement into the palatal region.

References:

Ladefoged and Maddieson (1996). The Sounds of the World's Languages. Oxford. Blackwell.

Poster Session Two: Thursday 7th November

Poster 2:

Velar coarticulation and tongue twisters in people who stutter 1: Fluent Utterances

Stefan A. Frisch, Alissa Belmont, & Nathan D. Maxfield

University of South Florida

Introduction

This poster will present data from across the lifespan for typical control participants and people who stutter. The task involved repeating velarvowel consonant combinations in a carrier sentence (e.g. Say a

<u>cope</u> again, following Wodzinski & Frisch 2005 ASA presentation) followed by a set of tongue twisters designed to induce velaralveolar errors (e.g. top cap cop tab, following Pouplier & Goldstein 2005). The goal of the study is to determine whether there is evidence for differences in the speech production process for people with stuttering even during fluent and otherwise normal sounding utterances.
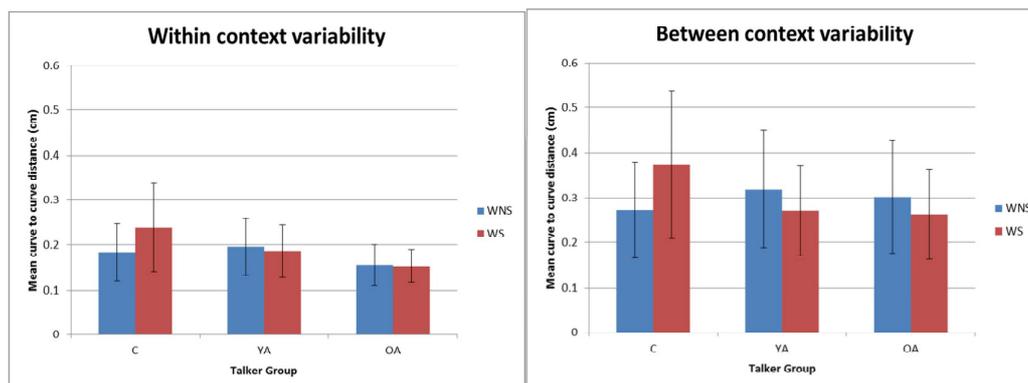
Analysis

Data for velarvowel coarticulation were analyzed using the Articulate Assistant Advanced software (Articulate Instruments, 2012) to create tongue traces. Tongue traces were then analyzed following the procedures for average of the nearest neighbor pointtopoint distance between curves (Zharkova & Hewlett 2009). Analysis of average nearest neighbor pointtopoint distance between tongue traces was conducted within a context as a measure of variability (e.g. a speaker's utterance of /ke/ compared to the speaker's other utterances of /ke/) and between contexts as a measure of coarticulation (e.g. a speaker's utterance of /ke/ compared to the speaker's other utterances of /ki/, /ko/, /ku/, etc).

Participants

30 speakers so far, 5 each from three age groups (children 812, young adults 1839, older adults 5065) and two populations (people with stuttering and typically fluent talkers).

Results

Children with stuttering had greater variability in productions both within and between contexts (less articulatory consistency and/or more holistic utterances). There were also trends toward older adults having less variability within context (more consistent repetitions) and adult stutterers having less variability between contexts (reduced coarticulatory variability).

Poster 4:

An investigation of tongue movement in patients with lingual hemiparalysis

Sina Koch (1), Tim Bressmann (2, 3)

(1)Rheinisch-Westfaelische Technische Hochschule Aachen
(2)Department of Speech-Language Pathology, University of Toronto
(3) *St. John's Rehabilitation Program, Sunnybrook Health Sciences Centre*

The functional consequences of a lingual hemiparalysis on tongue movement in speech production are not well understood. The present study used ultrasound imaging in the coronal plane to investigate lingual hemiparalysis. Ten stroke patients with lingual hemiparalysis and 6 normal speakers participated. The participants produced VCV combinations with the vowels [a, i, u] and the target consonants [t, k, s, ʃ]. The movement of the tongue in the coronal plane was traced in the ultrasound image. A series of Wilcoxon signed rank tests indicated that the paralyzed side of the tongue was

significantly lower than the unaffected side in the stroke patients. Further calculations regarding symmetry and grooving of the tongue are currently underway. The study provides first insights into the consequences of lingual hemiparalysis for tongue movement in speech.

Poster 6:

The articulation of Scottish Gaelic plain and palatalized consonants

Jae-Hyun Sung[1], Diana Archangeli[13], Ian Clayton[12], Daniel Brenner[1], Samuel Johnston[1], Michael Hammond[1], Andrew Carnie[1]

[1]*Department of linguistics, University of Arizona, USA*
[2]*Department of linguistics, University of Hong Kong, HK*
[3]*Deparment of English, Boise State University, USA*

Scottish Gaelic (Gàidhlig, henceforth SG) exhibits a rich system of consonant mutation, which is mostly governed by its morphology. For instance, the initial consonant [p] of a SG word *bàta* 'boat' changes to [v] when the word undergoes morphological inflection – e.g., *a bhàta* 'his boat', in which *bh* is pronounced as [v]. Although consonant mutation in SG is well documented (Macaulay (1992), Ladefoged et al. (1998), Gillies (2002), Stewart (2004)), there is only one experimental, quantitative study on this to date (Archangeli et al. (2011)). Our study is the first attempt to provide an empirical comparison of the articulation of unmutated vs. mutated consonants in SG using ultrasound imaging, focusing on plain (unmutated) vs. palatalized (mutated) consonants.

Palatalization appears as one of possible consonant mutation types in SG. Examples of palatalization are described in Table 1, in which (1a) and (1b) represent word-medial cases, and (1c) and (1d) represent word-final ones. Table 1 demonstrates that palatalization occurs in SG for different consonant types, in different syllabic positions, and different morphological contexts.

	Example	Effect	Gloss
a.	ar[d] - air[d]e	[t]→[tʲ]	high - higher/highes (comp./sup.)
b.	ò[g] - òi[g]e	[g]→[gʲ]	young - younger/youngest (comp./sup.)
c.	go[b] - gui[b]	[p]→[pʲ]	beak (nom. Sg.) - beak's (gen.sg.)
d.	dù[n] - dùi[n]	[n]→[nʲ]	fort (sg.) - forts (pl.)

Table 1: Palatalization in Scottish Gaelic

Our study investigates the articulation of plain and palatalized consonants in SG across different consonant types and syllabic positions. Experimental data was collected in Sabhal Mòr Ostaig, a college on the Isle of Skye, in which speakers of the Skye dialect are accessible. Whether plain vs. palatalized consonants differ in terms of articulation was determined based on differences in tongue contours – i.e., differences in the position of the tongue tip and back.

Preliminary results from 4 SG speakers show a clear sign of palatalizsation across different consonant types in palatalization environments (i.e., when morphologically conditioned, often preceded by an orthographic <i>), represented by higher tongue contours in the front region of tongue. While the articulatory distinction between plain and palatalized consonants is significant, different syllabic positions (i.e., word-initial vs. –medial vs. –final palatalization) do not yield any consistent trend in

terms of tongue gestures. Although they are based on a small sample of the population, the articulatory patterns of palatalization in different syllabic positions may not manifest as gestural differences.

The overall goal of this study is to look for the articulatory differences giving rise to the perceptual distinctions between plain and palatalized consonants. Furthermore, as part of the Arizona Scottish Gaelic project, the study will complement other components of the project such as statistical corpus analysis and psycholinguistic experimentation.

References

Archangeli, D., Berry, J., Hunt. N., Ji, S., Josephs, K., and Carnie, A. 2011. ATR in Scottish Gaelic tense sonorants: a preliminary report. In Carnie, A., editor, *Formal Approaches in Celtic Linguistics,* pages 283-306. Cambridge Scholars Publishing.

Gillies, W. 2002 Scottish Gaelic. In Ball, M. and Fife, J., editors, *The Celtic Languages,* pages 145-227. Routledge.

Ladefoged, P., Ladefoged, J., Turk, A., Hind, K., and Skilton, S. J. 1998. Phonetic structures of Scottish Gaelic. *Journal of International Phonetic Association,* 28:1-41

Macaulay, D. 1992. The Scottish Gaelic Language. In Macaulay, D., editor, *The Celtic Languages,* pages 137-248. Cambridge University Press.

Stewart, T. W. 2004. *Mutation as morphology: bases, stems, and shapes in Scottish Gaelic.* PhD thesis, The Ohio State University.

Poster 8:

An ultrasound analysis of the acquisition of rounded French vowels /y/ and /ø/ by native speakers of Brazilian Portugeuse

Giovana Ferreira-Gonçalves, Mirian Rose Brum-de-Oaula, Giulian Silva-Pinto

Universidade Federal de Pelotas

Key words: foreign language acquisition; front rounded vowels; French; ultrasound analysis

This paper discusses the acquisition of rounded French vowels /y/ and /Ø/ by native speakers of Brazilian Portuguese. When it comes to French as a Second Language, the acquisition of rounded vowels is hampered by the differences in the vowel systems of both the target language and the mother tongue (Brazilian Portuguese), which does not contain /y/ and /Ø/ vowels. Thus, the aim of this study is to describe the process of acquisition of rounded vowels by Brazilian adults who learn French in a university located in Southern Brazil. The corpus consists of data from 8 female students of different ages who are attending 2nd, 4th, 6th and 8th semesters. The subjects had to take a placement test, so that they could be divided according to their level of proficiency, namely Elementary, Intermediate or Advanced. Two other tests were carried out thereafter: (i) production test – where speech data was collected using a *Zoom H4N* digital voice recorder and a *Mindray DP6600* portable ultrasound machine; (ii) sound perception test, using the software *Praat 5.3.34*. The words were spoken separately, so that the ultrasound images could be best recorded. The collected data was analysed by Articulate Assistant Advanced (version 2.14) software. It suggests that there wasn't a complete acquisition of the front rounded vowel /y/ by students attending 1st, 3th and 5th semesters, since they would frequently replace it with [u] – the vowel that the subjects used more often – , [ju] and [i]. Furthermore, both the acoustic and articulatory analysis showed that the production of the [Ø] sound was closer to the native French speech than the close front rounded vowel [y], which was produced as [u], considering the movement patters of them. The results of the second test indicate that the improvements on sound perception take

place along with the improvement on the production of the rounded vowels at issue.

Poster 10:

An ultrasound study of oral and nasal central vowels in Brazilian Portuguese

Lilian C Kuhn Pereira a, Maria A F Svicero, Sandra Madu ureira, Zuleica Camargo

Integrated Acoustic A Analisys an Cognition Laboratory(LIAAC) – Pontifical Catho olic University of Sao Paulo (PUCSP), Sao Paulo, Brazil

Goal

The ultrasound technique is s a non-invasive way of investigating speech so ound productions and describing them. This s study aims at characterizing, by means of ultras sound techniques and acoustic analysis, the d differences in tongue position between the centraal oral and nasal vowels in syllable stressed d position and the central vowel in stressed and unstressed positions in Brazilian Portuguuese (BP). It is also concerned with approaching g its findings with the results of BP vowel desc criptions based on cine-radiographic data (Matta MMachado, 1981), acoustic analysis (Mendes, 2003), MRI (Demolin & Medeiros, 2006; Gregioo,2006), and US (Berti, 2010; Svicero, 2012).
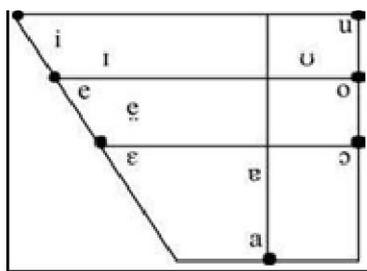
Figure 1: Oral monophthongs vvowels of BP

Oral vowels	Pre-stressed position (5)	a	e		i	o	u
	Stressed position (7)	a	ɛ	e	i	ɔ	o	u
	Post-stressed position (3)	a	i			u		

Figure 2: Oral vowels of BP in all three positions

Nasal Vowels	ẽ	ẽ	ĩ	õ	ũ

Figure 3: Nasal vowels of BP inn stressed positions

Methods

Speech samples from 2 female adult speakers of BP, with no speech, voice or auditory disorders, were collected. The corpora comprised eight non-sense words: /fapa/, /fapi/, /fapu/, /fipa/, /fupa/, /fãpa/, /fãpi/, /fãpu/ inserted in carrier sentences. The data were recorded at the Speech Science Research Laboratory at Queen Margaret University. A Concept M6 Digital Ultrasonic Diagnostic Imaging System, with a 120-degree transducer, was used to capture data. The ultrasound system, in

conjunction with the AAA software, captured 30 frames per second. Each recording session lasted for 30 minutes. The subjects performed a reading task and repeated the sentences, which were randomly presented four times. Simultaneous recordings of the ultrasound and the audio data, using the Articulate Assistant Advanced (AAA) software, were made. The speakers wore adjustable aluminum headsets during the whole recording session. This device, which was developed at QMU, is necessary to maintain the US probe in a fixed position under the chin. At the beginning of the recording session, as part of the preparation procedure, the subjects were asked to drink water and to read some words while biting a plastic device. In this study, the palate tracings were used to describe the subjects' articulatory movements and the Articulate Assistant Advanced software was used to annotate the spectrograms and to draw the splines of tongue contours on the ultrasound images. The AAA software produces the splines automatically and those which were not found to be well-drawn were done manually.

Analysis

Sounds whose tongue surfaces are flat and smoothly curved (e.g. lower vowels) provide the best images. On the other hand, the less visible images come from sounds as high vowels. So, depending on the data, it could be more or less difficult to identify the tongue shape, and consequently, the precision of spline´s drawing could be prejudiced. In this specific study, we dealt with /a/ (less difficult), /u/ (intermediary difficult) and /i/ (most difficult). In this study, the palate tracings were used to describe the subjects' articulatory movements and the Articulate Assistant Advanced software was used to annotate the spectrograms and to draw the splines of tongue contours on the ultrasound images. The AAA software produces the splines automatically and those which were not found to be well-drawn were done manually. Results of the splines and the acoustic analyses are discussed and considered in relation to the results of the other experimental (acoustic and MRI) works on BP vowels which were taken into account in this study. After that, in AAA software, we exported the data files for an excel table. There were some important values: duration (ms) of each vowel, and 42 values representing the distance (mm) from the original point for the curve for each grid of reflection.
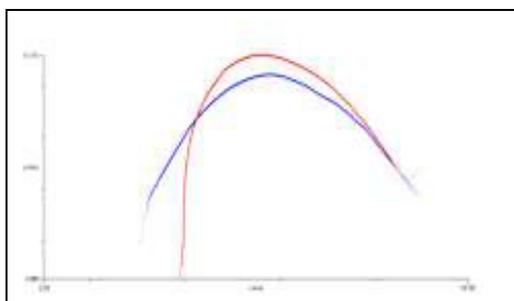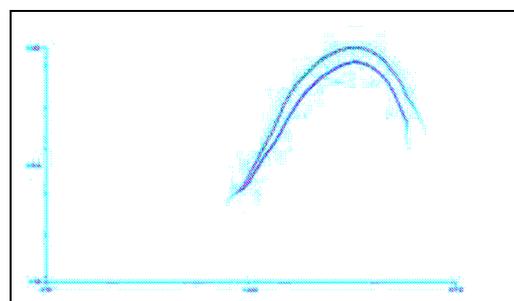

Figure 4: **Splines of vowels** /a/ **and** /ẽ/   Figure 5: Splines of vowels /i/ and / ĩ /
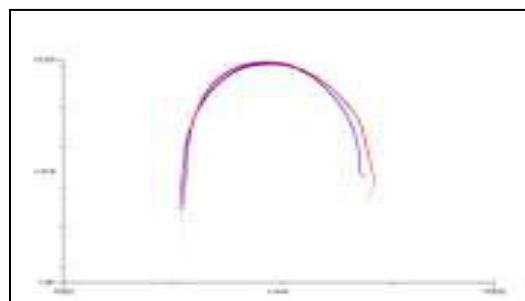

Figure 6: Splines of vowels /**u**/ and /**ũ**/

Results

Results of the splines and the acoustic analyses are discussed and considered in relation to the results of the other experimental (acoustic and MRI) works on BP vowels which were taken into account in this

study.

According to Svicero (2012), US image data revealed a better differentiation in the vertical axis the tongue body contour for BP seven oral vowels than horizontal axis. Comparing the three oral vowels, the present study had the same conclusion. In addition, for the comparison between each oral vowel with the nasal correspondent showed that both axes were relevant.

Conclusion

This US data preliminary study showed the US technique were relevant for language description and for clinical applications. Comparing to other techniques (MRI, Rx), ultrasound images may offer more advantages as being a non-invasive method. In this way, for future directions, the study will continue using hearing-impaired and/or speech disease data.

References:

BISOL, L., MAGALHÃES, J. S. Vowel reduction in Brazilian Portuguese. http://webs2002.uab.es/filologiacatalana/papi/ABSTRACTS/Abstract69-Affil.pdf, 2002.

MACHADO, M. T. M Etude Articulatoires et acoustique des voyelles nasales du Portugais du Rio de janeiro: Analyses Radiocinematographique, sonagraphique et oscilographique. Universite des Sciences Humaines de Strasbourg. Tese de doutorado, 1981.

MENDES, B.C.A. Estudo Fonético acústico das vogais do português brasileiro. São Paulo: Pontifícia Universidade Católica de São Paulo. Tese de doutorado, 2003.

DEMOLIN, D.; MEDEIROS, B.R. Vogais nasais do português brasileiro: um estudo de IRM. Revista da ABRALIN, 2006. V.5, n.1 e 2, p. 131-42.

GREGIO, F.N. Configuração do trato vocal supraglótico na produção das vogais do Português Brasileiro: dados de imagens de ressonância magnética. São Paulo: Pontifícia Universidade Católica de São Paulo. Dissertação de mestrado, 2006.

BERTI, L.C. Investigação da produção de fala a partir da ultrassonografia do movimento de língua. In: 18º Congresso Brasileiro de Fonoaudiologia: Exercício Profissional: Bases teóricas, avanço e realidade nacional. SBFa, Curitiba. 2010, v.1, p.661-1.

SVICERO, M.A.F. Caracterização Acústica e de imagens de ultrassonografia da vogais orais do Português Brasileiro. São Paulo: Pontifícia Universidade Católica de São Paulo. Dissertação de mestrado, 2012.

Poster 12:

Comparing the effect of pronunciation and ultrasound trainings to pronunciation training only for the improvement of the production of the French /y/-/u/ contrast by four Japanese learners of French

Tanja Kocjančič Antolík[1], Claire Pillot-Loiseau[1], Takeki Kamiyama[2, 1]

[1]Laboratoire de Phonétique et Phonologie (LPP) UMR 7018 CNRS) / Université Paris 3 Sorbonne Nouvelle
[2]Linguistique Anglaise Psycholinguistique (LAPS), EA 1569, Université Paris 8

Japanese learners of French commonly have difficulties producing perceptually recognisable French /y/ and /u/. The two sounds are articulatorily different from Japanese /u/, a high non-front vowel in the Tokyo variety with the tongue less retracted and lips less rounded than in French /u/ (Bothorel *et al.,* 1986; Uemura & Takada, 1990). The French /u/ produced by native speakers of Tokyo Japanese is typically perceived as /ø/ by native listeners of French (Kamiyama & Vaissière, 2009). French /u/ and /y/ are phonemically contrastive in Parisian French and are present in a number of minimal pairs. Being able to produce perceptually recognisable French /u/ is thus necessary for Japanese learners,

even more so because the more anterior realisations may overlap with French /y/. The two sounds are typically learned by way of perception, but the progress can be slow because of the lack of perceptual differentiation between them. For this reason, it was hypothesised that learners benefit from visual feedback of tongue position by avoiding relying only on perceptual route but rather addressing articulation directly.

In total, seven participants took part in the study. The first subject was a 42-year-old female French native, recorded for reference articulatory and acoustic data. The six others were adult female native speakers of Japanese and living in Paris at the time of the study. They all started learning French as adults, were intermediate level learners and, at the time of participation, they were all attending a 12-week French pronunciation course including training sessions in language lab. Four of the participants (experimental group) received three 45-minute training sessions in which ultrasound was used as a visual aid in achieving and controlling the tongue position of the target vowels. The training began with isolated vowels, progressed to non-words with different phonetic contexts (facilitating, neutral and difficult contexts), then on to real words and sentences. The exact protocol was adjusted to the abilities and preferences of each participant. Each of these four participants underwent ultrasound and audio recordings three times: one week before the first training session (pre-training), one week after the last training session (post-training) and two months after the post-training recording (follow-up). The remaining three participants did not receive any ultrasound training (control group) but were also recorded two times: at the beginning and at the end of the pronunciation course. The recorded corpus consisted of ten repetitions of (1) [y] and [u] in isolation, as well as [a], [i] and the Japanese [ɯ] (not recorded in pre-training), (2) alternation between [y] and [u], (3) disyllabic non-words CVCV where V is /y/ or /u/, and C is /p/, /t/ or /k/, (4) 28 real words and (5) four sentences (not recorded in pre-training). The French native speaker was recorded only once. Acoustic and articulatory analyses are under way. In this abstract, we focus on the articulatory data for two of the subjects in the experimental group, the two in the control group and for the native subject.
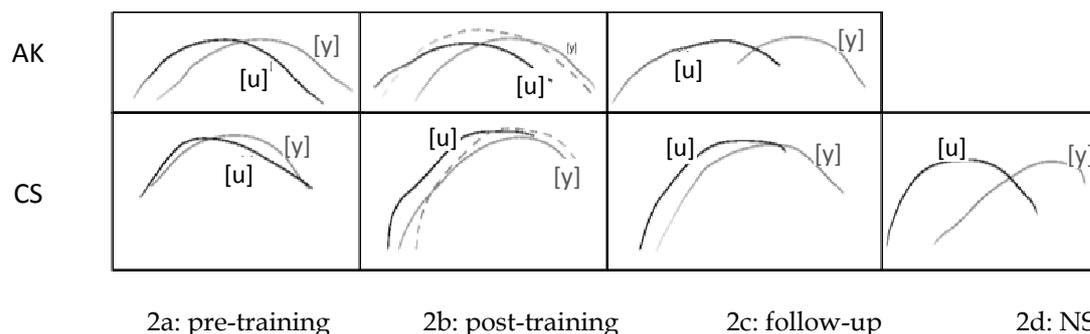

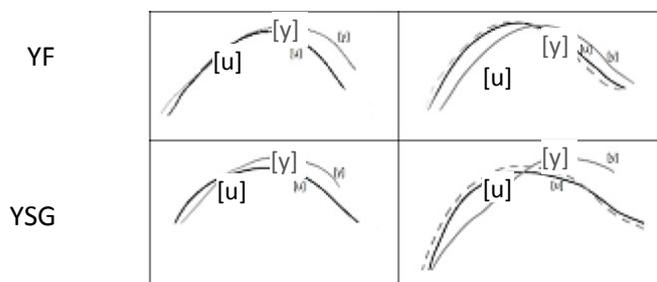| | 2a: pre-training | 2b: post-training | 2c: follow-up | 2d: NS |

Figure 1: Mean tongue position (10 repetitions) for [y] (solid grey curve), [u] (black curve) and [ɯ] (dashed grey curve) for AK (top) and CS (bottom) in the Experimental Group. From left to right: pre-training (2a), post-training (2b), follow-up (2c), native French speaker (2d). Front of the tongue is on the right side of each image.

The current analysis of isolated /y/ and /u/ of the two Japanese learners who received ultrasound training confirms the difficulties that Japanese learners of French have with the production of French /y/ and /u/ (Kamiyama and Vaissière 2009). It also shows some improvements both in acoustic and in articulatory (Figure 1) data in post-training: AK shows a clearer separation between the tongue contours for each of the two vowels in the post-training recording, and even a greater distinction in follow-up. CS showed a further posteriorisation of the tongue root for /u/ after the ultrasound trainings (Figure 1), while the tongue shape is similar between the two recordings for both control learners: their /u/ is similar to the Japanese [ɯ] after traditional pronunciation lessons (figure 2).

The four Japanese learners who received lessons with ultrasound reported that these sessions were enjoyable and effective for both /y/ and /u/. They said that the image helped them to better control the position of their tongue. 3 of 4 speakers easily understood this image, which helped them a lot. All of them would take some lessons with ultrasound to better articulate other French sounds.

This study is a first step. Further analysis of the available data will allow inspecting articulatory improvements in more varied contexts (words vs. nonwords, mono- vs. disyllabic words, isolated vowels vs. words vs. sentences) as well as in more quantitative details (tongue height and curvature, tongue curvature position, Ménard *et al.* 2012; Dorsum Excursion Index DEI, Zharkova 2013, among others parameters). The method will also be useful for the analysis of other kinds of productions such as the singing voice.

Figure 2: Mean tongue position (10 repetitions) for [y] (solid grey curve), [u] (black curve) and [ɯ] (dashed grey curve) for YF (top) and YSG (bottom) in the Control Group.

References

Bothorel A., Simon P., Wioland F., Zerling J.P. (1986). *Cinéradiographie des voyelles et consonnes du français*. Strasbourg, Publications de l'Institut de Phonétique de Strasbourg.

Kamiyama T., Vaissière, J. (2009). Perception and production of French close and close-mid rounded vowels by Japanese-speaking learners, *Acquisition et interaction en langue étrangère Aile... Lia* 2, 9-41.

Ménard, L., Aubin, J., Thibeault, M., Richard, G. (2012). Measuring Tongue Shapes and Positions with Ultrasound Imaging: A Validation Experiment Using an Articulatory Model, Folia Phoniatrica et Logopaedica, 64, 64-72.

Uemura Y., Takada M. (Kokuritsu Kokugo Kenkyûjo [National Institute for the Japanese Language]) (1990). *Nihongo no boin, shiin, onsetsu: chouon undou no jikken-onseigakuteki kenkyuu* [Vowels, consonants and syllables in Japanese: experimental phonetics study in articulatory movements]. Tokyo, Shûei shuppan.

Zharkova, N. (2013). A normative-speaker validation study of two indices developed to quantify tongue dorsum activity from midsagittal tongue shapes, *Clinical Linguistics & Phonetics*, 27(6-7), 484-496.

Poster 14:

An Ultrasound Analysis of Tongue Shape for the Japanese Nasal Mora

Ai Mizoguchi and Douglas H. Whalen

City University New York

The nasal mora in Japanese occurs at the end of words and takes its place of articulation from the following segment (Amanuma, Otsubo, & Mizutani, 1978; Nakajo, 1989; Vance, 2008). In utterance-final position, the nasal mora is variously transcribed as alveolar, velar or uvular in place, but it may simply lack a complete constriction and thus not be any of these places (Vance, 2008: 101). In that case, it may be that the nasal mora does not assimilate to the place of a following consonant; it may simply be coproduced with it, and the result is dominated by the full(er) closure of the conditioning consonant.

The present pilot study examines two speakers of Tokyo Japanese to determine whether their tongue shape for the nasal mora remained relatively unchanged despite apparent assimilation to following consonants. For those consonants that also use the tongue (lamino-alveolars, laminoalveopalatals, the apico-alveolar flap, and the velars), some modification can be expected. However, we hypothesize that the shape will differ between those consonants with and without a preceding nasal mora.

The participants will be instructed to speak 10 repetitions of 23 words including one utterance-final /N/, 11 contextual allophones of /N/, and 11 cases of those same conditioning consonants without a preceding nasal mora. The tongue will be imaged via ultrasound and placed into head-relative space with a specialized system. We expect to find that the tongue shape for utterance-final nasal moras is relatively back but without contact with the upper vocal tract. We further expect to see that shape or remnants of it in the allophonic shapes before consonants.

References:

Amanuma, Y., Otsubo, K., & Mizutani, O. (1978). *Nihongo onseigaku*. Tokyo: Kuroshio Shuppan.
Nakajo, O. (1989). *Nihongo no onin to akusento (Dai 1-han. ed.)*. Tokyo: Keiso Shobo
Vance, T. J. (2008). *The sounds of Japanese*. Cambridge, UK ; New York: Cambridge University Press.

Poster 16:

Many-to-one articulatory-acoustic mappings in Canadian French vowels

Will Dalton and Jeff Mielke

University of Ottowa

We report an ultrasound investigation of two cases of many-to-one articulatory-acoustic mapping in Canadian French. The first is an allophony pattern found in high vowels, whereby non-lengthened /i y u/ in closed syllables are realized as [ɪ ʏ ʊ]. While this has traditionally been treated as a tense-lax alternation, the main acoustic feature of the lax allophones (higher F1) can be achieved either by tongue root retraction or by tongue body lowering. Our articulatory results show inter-speaker variability between the two strategies. The second case involves a change in progress in which mid front rounded vowels /ø œ œ̃ / are produced with a rhotic perceptual quality, much like English [ɹ]. Whereas previous reports (e.g., Dumas 1972, 100) have described the innovative variant as retroflex, our results show that most speakers produce it with a bunched tongue shape, and one produces it with retroflexion.

Ultrasound imaging was used to produce mid-sagittal tongue video for 23 native speakers of Canadian French (including Laurentian and Acadian varieties) as they pronounced words containing / i y u ø œ œ̃ / in a carrier phrase, resulting in about 200 tokens of the target vowels for each participant. Smoothing-Spline ANOVA with polar coordinates was used to compare tense and lax allophones of high vowels, and to compare potentially rhotic mid front rounded vowels with non-rhotic mid front unrounded vowels.
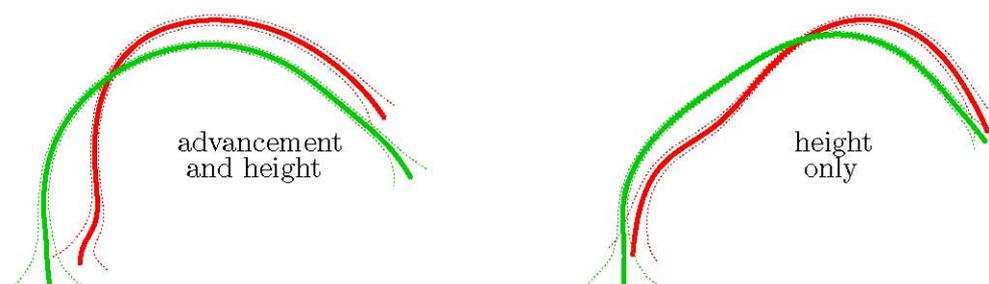

Figure 1: Tongue contours for two speakers' [i]~[ɪ] alternations

The SSANOVA comparisons of corresponding tense and lax high vowels revealed that the speakers are evenly split between two categories: those who realize the tense-lax difference as a difference in tongue

body height, and those who realize both height and tongue root advancement (Figure 1). No speakers realize the distinction with tongue root advancement alone. This has phonological implications, considering that Canadian French also exhibits a phenomenon typically described as regressive ATR harmony (Poliquin, 2006), triggered by the lax allophones.

Articulatory imaging of rhotic vowels reveals that bunched tongue shapes are used to produce all moderately rhotic vowels and most extremely rhotic vowels, but that one extremely rhotic speaker uses two distinct retroflex tongue shapes to produce different rhotic vowels. Figure 2 shows examples of tongue shapes for two rhotic speakers. A corpus study of Gatineau French (Mielke, accepted) showed that the development of rhotic vowels has been gradual, and appears not to be direct result of contact with English (the change has been in progress for several decades and the vowels involved only recently became perceptually similar to English /ɝ/).

A further curiosity of this change in progress is that the shift from front-rounded vowel to rhotic vowel appears to involve a change in cavity affliation of the second and third formants. Zhou et al. (2008) provide evidence from MRI and acoustic modeling to show that F1, F3, F4, and F5 are all back cavity resonances for bunched and retroflex English /ɹ/, and F2 is a resonance of the front cavity. This contrasts with front rounded vowels, where F3 is typically a front cavity resonance, and thereby more sensitive to lip rounding than the back cavity resonance F2 (e.g., Ladefoged 1996). Figure 3 shows representative spectrograms for *beurr*e spoken by a rhotic and a non-rhotic speaker. Arrows identify F3 during the target vowel. The rhotic speaker shows lowering of F3 into F2's range, whereas the non-rhotic speaker shows comparatively even spacing between the formants. The spectrogram for rhotic *beurr*e also shows that the uvular /ʁ/ is still present following the rhotic vowel (which is diphthongized), and the the low F3 target is achieved before the end of the vowel.
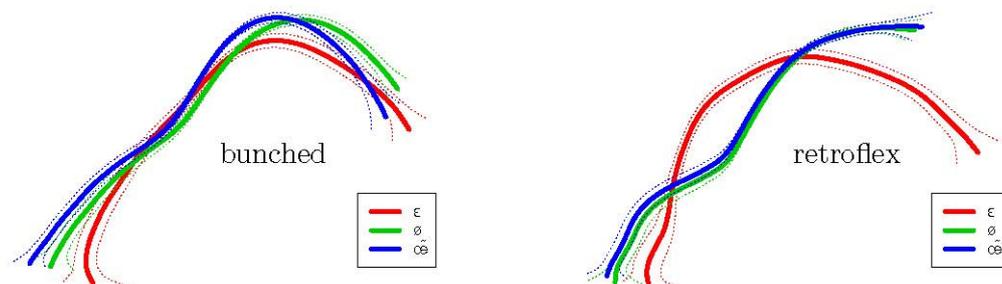

Figure 2: Tongue contours for two speakers' rhotic vowels, with non-rhotic vowel [ɛ] for reference.
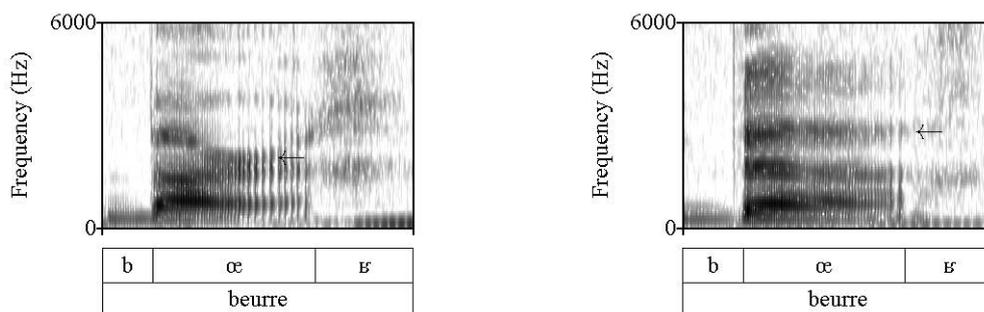

Figure 3: Representative spectrograms (left: rhotic; right: non-rhotic). Arrows identify F3.

We will show the consequences of the height/advancements trade-off for the leftward spread of laxing often referred to as Canadian French vowel harmony (which our data show is not categorical), and the relationship between the gradual advancement of vowel rhoticity and the categorical nature of bunched/retroflex variation and formant/cavity affliation.

References

Dumas, Denis. 1972. Le français populaire de Montréal: description phonologique. Université de Montréal M.A. thesis.

Ladefoged, Peter. 1996. *Element*s *of acousti*c *phonetics*. Chicago: University of Chicago Press.

Mielke, Jeff. accepted. Ultrasound and corpus study of a change from below: Vowel rhoticity in

Canadian French. In *Penn Working Papers in Linguistics 19.2: Papers from NWAV 41* .

Poliquin, Gabriel. 2006. Canadian French vowel harmony. Doctoral Dissertation, Harvard University.

Zhou, Xinhui, Carol Y. Espy-Wilson, Suzanne Boyce, Mark Tiede, Christy Holland, and Ann Choe. 2008. A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/. *Journal of the Acoustical Society of America* 123:4466–4481.

Poster 18:

Factors affecting the articulatory and acoustic characteristics of vowels in clear speech

Jae Yung Song
University of Wisconsin-Milwaukee

When we speak, we are constantly modifying our speech to fit the environment and the listener. For example, when we talk to people with hearing loss, we slow down our speech and articulate individual sounds more carefully to make it easier for the listener to perceive words. This type of unique speech register is known as 'clear speech.' While there has been much research done on the acoustic properties of clear speech, relatively little is known about the articulatory basis of the previous acoustic findings. Furthermore, because many studies have focused on a single factor, only limited information is available on how various factors interact to affect the speaker's production of clear speech. Thus, the purpose of this study was twofold: to examine the specific articulatory processes of clear speech for individual vowels using ultrasound, and to provide a more comprehensive investigation of the factors affecting the production of clear speech. In particular, we investigated how the type of listener (hard of hearing vs. normal hearing) and the lexical properties of words (frequency and phonological neighborhood density) interact.

To address the issues, 18 monolingual speakers of American English (10 males, 8 females) participated in the experiment where they were asked to read a set of words (varying in word frequency (F) and phonological neighborhood density (D)) once to a hard of hearing listener and once to a normal hearing listener presented in video. Each speaker read 96 words in total: 2 listeners (hard of hearing vs. normal hearing) × 4 lexical properties (HighF-HighD, HighF-LowD, LowF-HighD, LowF-LowD) × 2 vowels (/æ, ɑ/) × 3 words per vowel × 2 repetitions of each word. The audio was collected using a Shure KSM137 microphone, and the video of tongue movements was collected using a Sonosite 180 Plus ultrasound machine and a C11/7-4 MHz 11-mm broadband curved array transducer.

First, we examined how two acoustic parameters, vowel duration and vowel space, vary as a function of listener type, word frequency, and neighborhood density. Three-way repeated measures of ANOVA showed that, as expected, there was the main effect of the listener type; that is, more clear speech (longer vowel duration and expanded vowel space) was directed to the listener who was hard of hearing. The main effects of word frequency and neighborhood density were also significant, with more clear speech for low frequency words and low density words. Overall, there was lack of evidence of interaction between the listener type and lexical properties of words. One exception was the listener type and density interaction effect on vowel duration; that is, speakers made extra efforts to lengthen low density words compared to high density words, in particular, for a hard of hearing listener.

Second, we investigated the articulatory details of clear speech, focusing on four parameters quantifying the shape and position of tongue contours: asymmetry, curvature, and height and frontness of the highest point of the tongue (Aubin & Ménard, 2006; Ménard, Aubin, Thibeault, & Richard, 2012). Overall, the ultrasound data confirmed that individual vowels were more separated and less variable in clear speech (Figure 1). For the low front vowel /æ/, clarity was mainly achieved by adjusting the tongue height and curvature; that is, the tongue was lower and flatter for /æ/, exhibiting its more 'canonical' form in clear speech.

For the low back vowel /ɑ/, the primary parameter changed was the tongue frontness and asymmetry, with the position and mass of the tongue moving toward the back in clear speech. Taken together, our

ultrasound results suggest that clear speech vowel adaptations are dynamic with both vowel-general and specific transformations.
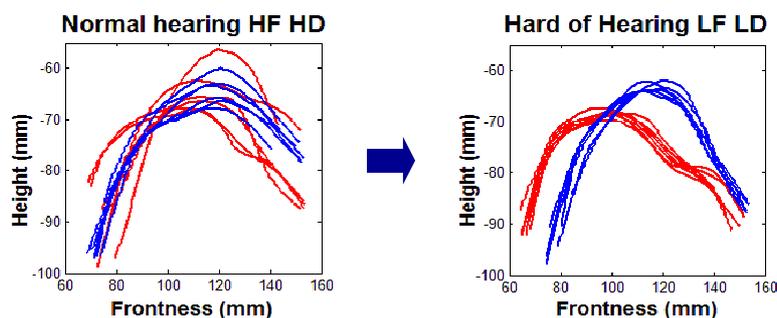

Figure 1. Representative changes in tongue contours from the condition where the participants least produced clear speech (i.e., normal hearing-HighF-HighD) to the condition where the participants most produced clear speech (i.e., hard of hearing-LowF-LowD) for the two vowels examined: /ɑ/ (red lines), /æ/ (blue lines). Individual repetitions of words from one speaker are shown on the plots.

References:

Aubin, J., & Ménard, L. (2006). Compensation for a labial perturbation: An acoustic and articulatory study of child and adult French speakers. *Proceedings of the 7th International Seminar on Speech Production*, Ubatuba, Brazil, 209-216.

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

Poster 20:

A New Head-Probe Stabilization Device for synchronized Ultrasound and Electromagnetic Articulography recordings

F. Sigona[1], A. Stella[1], B. Gili Fivela[1], F. Montagna[2], A. Maffezzoli[2], A. Wrench[3] and M. Grimaldi[1]

[1]*Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL) – University of Salento (Lecce – Italy)*
[2]*Dipartimento di Ingegneria dell'Innovazione – University of Salento (Lecce – Italy)*
[3]*Clinical Audiology, Speech and Language Research Centre - Queen Margaret University (Edinburgh)*

In the few last decades, the technology development in both imaging and position tracking techniques allowed a refined study of articulatory correlates of speech production. Ultrasound waves are largely used to investigate the tongue motion during speech (Ultrasound Tongue Imaging, UTI), due to the acoustic properties of the tongue muscles, especially at the interface between the tongue dorsum and the upper air layer. In the panorama of the position tracking techniques stands the electromagnetic articulography (EMA), which exploits the physical properties of electromagnetic induction in order to track the position of the articulators in time (cf. [1], [3], [4], [5], [7], [8]).

The above mentioned technologies have complementary characteristics: e.g., EMA has high time resolution and can track the motion of articulators outside the internal of the mouth, but it supports a limited number of recording channels, whereas UTI is non-invasive, can provide a full profile of tongue dorsum (and sometime of the tongue root too), but it has a sensibly lower sampling rate. For this reason, their simultaneous use constitutes an added value to reach a deep understanding of the articulatory correlates of speech production, e.g. to put in relation the morphology of the tongue with the articulatory landmarks on the tongue surface or other oral articulators for a deeper understanding

of muscle structures coordination. In such cases, some further care is necessary. UTI typically requires a head-probe stabilization device in order to avoid misalignment during data collection that could lead to image a different tongue profile than the desired one. Various systems have been proposed and successfully used (see, among others, [2], [3], [6], [7], [10]), but none of them is adequate to support a synchronized EMA recording, due to the magnetic field distortion induced by metallic masses which would be introduced into the EMA measurement volume.

In order to overcome this problem, we designed and manufactured a new head-probe stabilization headset (Figure 1), inspired by [3] and [2], replacing the metal structure with a polycarbonate one, relying on the well-known physical properties of this material that does not interact with magnetic fields, thus not affecting at all the EMA sensors position tracking. This headset, that has a helmet form, has been tested in various speech production tasks during simultaneous speech/UTI/EMA recording sessions. In this contribution, after describing the synchronization system, we will present and discuss a series of tests performed by means of a Toshiba Aplio XV ultrasound equipment and both the AG500 and the new AG501 electromagnetic articulographs by Carstens Medizinelektronik GmbH. Provided that the headset is absolutely transparent to the EMA operation, the aim of the tests is to show that the helmet and the ultrasound probe do not significantly move over the subject, resulting in reliable ultrasound tongue images, and that these are well time-aligned to speech and EMA data.

Figure 1: The polycarbonate helmet

In order to test the performance of the synchronized system, we built a series of 10 pseudo-words composed by 4 syllables and stressed on the second one (Table 1). The words differ in the primary articulators employed for the production of the consonantal segments: i.e., constrictions are produced with different parts of the tongue and/or different values of the dynamical parameters of the articulatory movements involved. Furthermore, consonants are inserted in the vocalic cycles /i-a-i/ and /o-a-o/, in order to detect different coarticulation effects due to the influences of the vocalic context on the consonantal movements.

Pseudo-words	Primary consonantal articulator
/mi.'ma.mi.ma/, /mo.'ma.mo.ma/	lips, tongue actively involved only for the vocalic cycles
/ni.'na.ni.na/, /no.'na.no.na/	tongue tip, single constriction
/ri.'ra.ri.ra/, /ro.'ra.ro.ra/	tongue tip, repeated constrictions
/li.'la.li.la/, /lo.'la.lo.la/	tongue tip, more retracted than [n] and [r]
/gi.'ga.gi.ga/, /go.'ga.go.ga/	tongue root

Table 1: pseudo-words exploited to test the performances of the helmet and the synchronized UTI/EMA system.

The corpus used in this experiment is composed by target words produced in isolation and target words in YN questions (here only the former were discussed). In particular:

– Target words produced in isolation: the subject was asked to produce 3 consequent repetitions of each word. They were also asked to produce them with a contrastive intonation, in order to maximize the articulatory effort employed for the production of the stressed syllable.

– Target words in YN questions: each word was inserted in the carrier sentence "Hai visto X ?" ("Did you see X ?"), where X is the target word. The subject was asked to produce each question one time with a surprise intonation, which should be useful to amplify the head movements during the spoken production.

The stimuli are presented in random order using the *Linguometer* software [11] and the entire corpus was produced 7 times by the subject. We recorded the productions of one subject, collecting 70 triplets of words in isolation and 70 questions per subject.
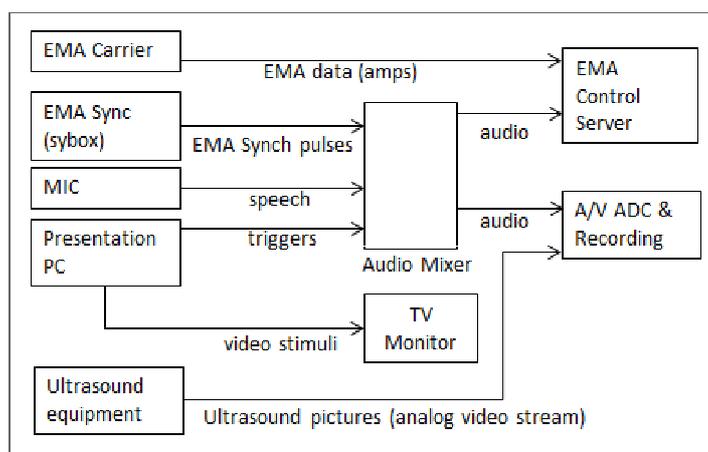
Figure 2: synchronized presentation and recording block diagram.

For each stimulus, the presentation equipment issues a couple of trigger pulses (begin/end stimulus) which are mixed to the speech signal and to the AG50x synchronization signal (Sybox), eventually input to the soundcard of the AG50x control server. The same speech+triggers signal is also input to an audio/video grabber together with the continuous PAL S-Video signal (25 fps) coming from the Ultrasound equipment (Figure 2). After the end of the recording session, the AG50x "calcpos" software routine can convert the EMA recorded data (250 sample/sec for AG501, 200 sample/sec for AG500) into articulator position tracking P(x,y,z,t). At this point, several *Linguometer* software routines can perform a first coarse time-alignment of speech/EMA/UTI data, based on the triggers timing on the audio traces, and then a second finer cross-correlation-based time-alignment and segmentation. At the end of this semiautomatic process, for each presented stimulus a speech/EMA/UTI data set is ready for whatever subsequent feature extraction (including tongue contour tracing) and analysis.EMA receiving coils were placed over the tongue at the intersection with the mid-sagittal plane (4), over upper and lower incisors (2), upper and lower lips (2), behind the ears (2), over the helmet (4), over the Ultrasound probe (1). The helmet was holding an ultrasound convex probe under the chin, in order to image a mid-sagittal section of the tongue surface.

Regarding the stability, the results showed that the new polycarbonate helmet can be adjusted to fit to the subject head and to assure the correct position of the US probe to the jaw, allowing an adequate freedom of movements of the oral articulators and of the head. In order to get a quantitative estimate of the goodness of blocking, the mutual Euclidean distance between the sensors over the probe, over the helmet and those placed behind the ears and over the upper teeth were considered. An ideal setup would require that each sensor position should not change relatively to other sensors, even if the subject produces speech in particular emphatic context, e.g. surprise intonation, which could stress the stability of the helmet to the head. In our study we found that the sensor distances were constants over time, they were varying no more than 5 mm in the worst case while the subject was naturally

producing the required corpus. This implies that the US probe was not moved so that the UTI data resulted perturbed.
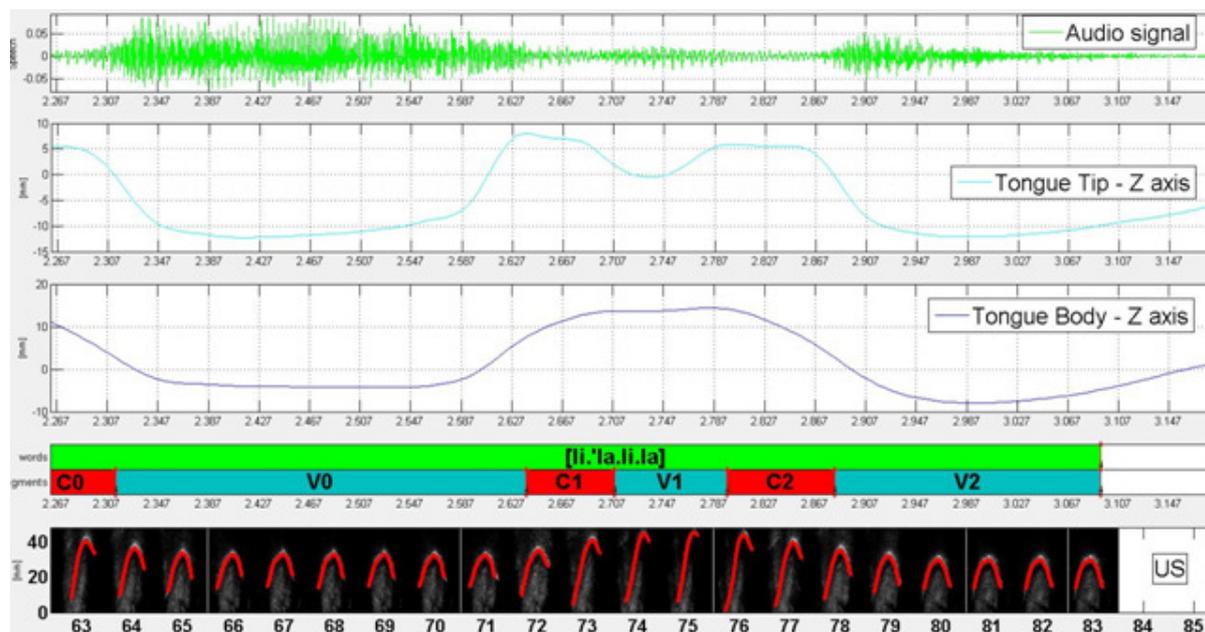

Figure 3: synchronized UTI/EMA signals of the pseudo-word [li.'la.li.la] as produced in isolation. From top to bottom: audio signal; tongue tip position track on Z axis; tongue body position track on Z axis; Praat textgrid, with the indication of the acustic duration of the segments; sequence of frames acquired by the UTI system. The plots are referred to the production in the time interval of the last [a-i-a] vocalic cycle of the pseudo word.

Data acquired with the synchronized UTI/EMA are particularly useful in the study of the tongue position with respect to its morphology and shape, in particular for those parts of the tongue from which the articulatory information is lost for intrinsic limits of the single technologies. Indeed, it is largely known that UTI system can't display the image of the tongue when it is extended towards the lips, because of the presence of the bones of the chin between the tongue apex and the UTI probe. For example, in producing apical consonants such [l] or [n], the articulatory information of the tongue tip gesture, and in particular the duration of the target attainment phase of the gesture, is unrecoverable from the UTI contours. This is shown in Figure 3, where the EMA tracks and the UTI frames of tonic, post-tonic, and final syllable of the pseudo-word [li.'la.li.la] are synchronized over time (cf. caption of the figure for tiers descriptions). Looking at the US frames interval 75-79, which displays the shape of the tongue in correspondence of the maximum constriction of the tongue tip for the production of the post-tonic consonant [l], the tongue apex is completely obscured by the chin and its position in the oral cavity is no more detectable. However, in the tongue tip position track as given by EMA (second tier in Figure 3), this information can still be recovered and can be directly correlated to the shape of the tongue, since both systems are synchronized in time. In this way no articulatory information about the gestures is lost. This advantage is crucial and useful for a deeper understanding of the coordination mechanisms of oral articulators and their relation to tongue muscle structures. Overall, the findings, according to our hypotheis, showed that the polycarbonate head-probe stabilization headset allows the successfully joint (time-aligned) EMA+UTI data recording without perturbations of the probe acquisition process and the EMA magnetic fields.

References

[1] Kaburagi, T., Wakamiya, K., and Honda, M. (2005), Three-dimensional electromagnetic articulography: A measurement principle, *Journal of Acoustical Society of America*. 118, 428–443.

[2] Miller, A.L. and Finch, K.B., (2011) Corrected High–Frame Rate Anchored Ultrasound With Software Alignment, *J. Speech Lang Hear Res*, 54: 471 - 486.

[3] 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.

[4] Stella, M., Stella, A., Sigona, F., Bernardini, P., Grimaldi, M., Gili Fivela, B., (2013), Electromagnetic Articulography with AG500 and AG501, in *14th Annual Conference of the International Speech Communication Association*, Lyon, France, 25-29 August 2013: 1316-1320.

[5] Stella, M., Bernardini, P., Sigona, F., Stella, A., Grimaldi, M., Gili Fivela, B., (2012) Numerical instabilities and three-dimensional electromagnetic articulography, in *Journal of Acoustical Society of America*, 132 (6).

[6] Stone, M., (2005), A Guide to Analyzing Tongue Motion from Ultrasound Images, *Clinical Linguistics and Phonetics,* 19, (6-7), 455-502

[7] Whalen, D.H., Iskarous, K., Tiede, M.K., Ostry, D.J., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E., and Hailey, D.S., (2005) The Haskins Optically Corrected Ultrasound System (HOCUS), *Journal of Speech Lang Hear Res.*; 48: 543 - 553.

[8] Yunusova, Y., Green, J. R., and Mefferd, A. (2009). "Accuracy assessment for AG500, electromagnetic articulograph," J. Speech Lang. Hear. Res. 52, 547–555.

[9] Zierdt, A., Hoole, P., and Tillmann, H. G. (1999), Development of a system for three dimensional fleshpoint measurement of speech movements, in *Proceedings of the XIV International Congress on Phonetic Sciences*, edited by J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Baile, San Francisco, CA, Vol. 1, pp. 73–76.

[10] Hueber, T., Benaroya, E.L., Chollet, G., Denby, B., Dreyfus, G., Stone, M., (2009), Development of a Silent Speech Interface Driven by Ultrasound and Optical Images of the Tongue and Lips, Speech Communication, 52(4), pp. 288-300.

[11] Grimaldi, M., Gili Fivela, B., Sigona, F., Tavella, M., Fitzpatrick, P., Craighero, L., Fadiga, L., et al. (2008). New technologies for simultaneous acquisition of speech articulatory data: ultrasound, 3D articulograph and electroglottograph, in *Atti del Convegno Internazionale LangTech*, pp. 81–85.

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

Poster 22:

An ANN approach to the recognition

of UTI-based Italian words

Alessandro Vietti, Lorenzo Spreafico, Vittorio Anselmi

Free University of Bozen-Bolzano

This paper presents the framework for a neural network aimed at recognizing ultrasound data from a limited set of words, to be included in a silent speech interface (Denby, Stone 2004; Hueber et al. 2007). On the whole the problems of characterization of a set of 20 words in Italian and their automatic identification will be presented and discussed. The corpus is based on ten repetitions of a word list as produced by three native speakers of Italian. For data collection, we used the Articulate Instruments multichannel acquisition system called Articulate Assistant Advanced™ (AAA) (Articulate Instruments 2011). Basically, we selected minimal pairs displaying a contrast in place (/ˈpattsa/ ~ /ˈtattsa/), manner of articulation (/ˈlana/ ~ /ˈtana/), and tongue profile (as a significant dimension of phonological contrast) in order to evaluate the different performance of our recognition system.

We provide a first proposal of a neural network model (Chicocki, Unbehauen 1993; Berry 2012) that can recognize lexical items from articulatory data only. The different stages in the construction of the model are described and discussed: data acquisition, articulatory signal filtering, signal characterization, the training of the neurons.

References

Berry, J. (2012). Machine Learning Methods for Articulatory Data. Ph.D. dissertation, University of Arizona.

Cichocki, A. and Unbehauen, R. (1993). Neural networks for optimization and signal processing. New York: Wiley.

Denby, B. and M. Stone (2004). Speech synthesis from real time ultrasound images of the tongue. In Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference, volume 1, pp. I-685. IEEE.

Hueber, T., G. Aversano, G. Chollet, B. Denby, G. Dreyfus, Y. Oussar, P. Roussel, and M. Stone (2007). EigenTongue feature extraction for an ultrasound-based silent speech interface. In Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on, volume 1. IEEE.

Poster 24:

Enhancing Measures of Reaction Time with UTI Data

Sonja Schaeffler and James M. Scobbie

CASL, Queen Margaret University, Edinburgh, UK

This paper presents an explorative study aimed at developing a suitable methodology to reliably determine the very first onset of target-related tongue movement for the purpose of verbal reaction time experiments.

In psycholinguistic experiments, reaction time is usually determined via "voice key", a device which is triggered automatically as soon as the sound pressure reaches a pre-defined level. However, what is detectable as the onset of acoustic output is only one stage in the speech production process. Before anything becomes audible, the articulators (i.e. the tongue, the lips, the jaw, etc.) have already moved into place. This movement shows the motor plan for the response has been put into action and is an earlier observable response than audible speech. Our data shows that in a standard picture naming task with a mean acoustic reaction time of 851 ms (SD 251 ms), observed articulatory reaction time occurs approximately 180 ms earlier.

What can make finding the exact onset of target-related tongue movement difficult is determining the boundary between that onset and some directly preceding, non-linguistic pre-speech behaviour. Before the articulators begin to accelerate towards their first stimulus-defined targets and the moment of noise generation, they may move from a static (closed or open) position, whatever the speaker's subconsciously chosen start position happens to be for that token. Alternatively, and this is more complex analytically, acceleration towards a linguistic target may originate from a moving position.

We will discuss the reliability of tongue onset measures in the light of these possible variations, and across a variety of different phonemic targets, and also discuss advantages and disadvantages in comparison with similar measurements in recordings of the lips.

Poster 26:

Ultrax: Vocal Tract Imaging for Speech Therapy and Research

Korin Richmond[1], Steve Renals[1], Joanne Cleland[2], Jim Scobbie[2], Alan Wrench[3]

[1]Centre for Speech Technology Research, Informatics, University of Edinburgh, UK
[2]Clinical Audiology, Speech and Language Research Centre, Queen Margaret University, UK
[3]Articulate Instruments Ltd, Edinburgh, UK

Background and motivation

Speech sound disorders are the most common communication impairment in childhood, affecting 1 in 15 children in the UK, for example. Sufferers have difficulty producing one or more speech sounds of their native language. This can affect a child's confidence, hindering communication with peers and teachers, and so harming learning as well as social interaction and development. Current speech therapy interventions mainly rely on the child's auditory skills, as little technology is available to help diagnosis and therapy. The Ultrax project (http://www.ultrax-speech.org) is a three-year project, funded by the EPSRC Healthcare Partnerships programme, which aims to use ultrasound to provide real-time visual feedback of lingual articulation. The ultimate aim is to exploit ultrasound to track tongue movement, while a simplified, diagrammatic vocal tract is displayed to the user as a realtime animation (see Fig. 1).

(a) Ultrasound speech therapy (note probe stabilisation mount)

(b) Raw ultrasound display

(c) Goal: a simplified display
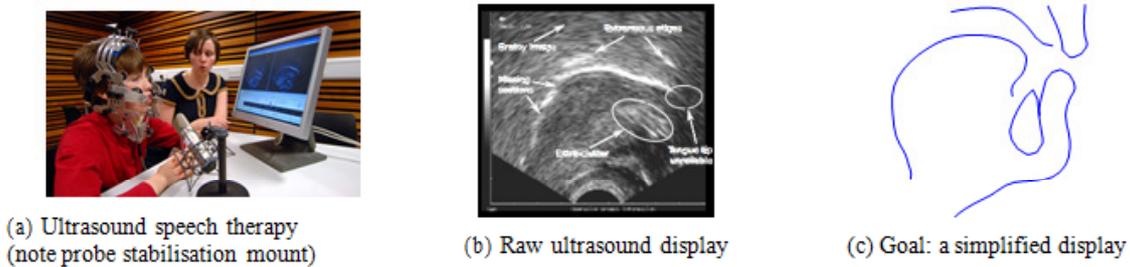
Figure 1: *a) Ultrasound can give visual feedback of lingual articulation for speech therapy. b) However, the native display has potential drawbacks for this purpose. c)* Ultrax *aims to develop a simple 2D animated display, driven by ultrasound.*

Approach and data

A machine learning approach is employed in conjunction with MRI and ultrasound imaging data. MRI is used to obtain a wide variety of vocal tract (VT) shapes (hand-labelled contours) for multiple subjects. This corpus is used to train a dimensionality-reducing VT shape model. The dimensionality reduction methods evaluated for this purpose include principal component analysis (PCA) and Laplacian eigenmaps [1].

Siemens Verio 3T scanner was used (1.7x1.7mm voxels, 52ms echo time, 400ms repetition time(=2.5fps), with an FOMRI-III optical mic for audio). Matching ultrasound data has been collected from the same subjects using an Ultrasonix RP system (B mode, 10mm microconvex probe at 5MHz), and similarly from around 90 typically-developing children performing a range of speech tasks. The control parameters of the above VT shape model are then used as the latent space representation of a Kalman filter variant [2], which in turn tracks the tongue in the ultrasound image sequence and animates the simplified VT display.

MRI data has been obtained from a total of 12 adult phoneticians, each producing 52 sound shapes. A Beyond speech therapy alone, the methods developed may also facilitate the collection of human articulography imaging data for other purposes. These include research in the fields of speech production and technology, for example silent speech interfaces (e.g. for laryngectomy patients), foreign language training, automatic speech recognition, text-to-speech synthesis and visual speech animation.
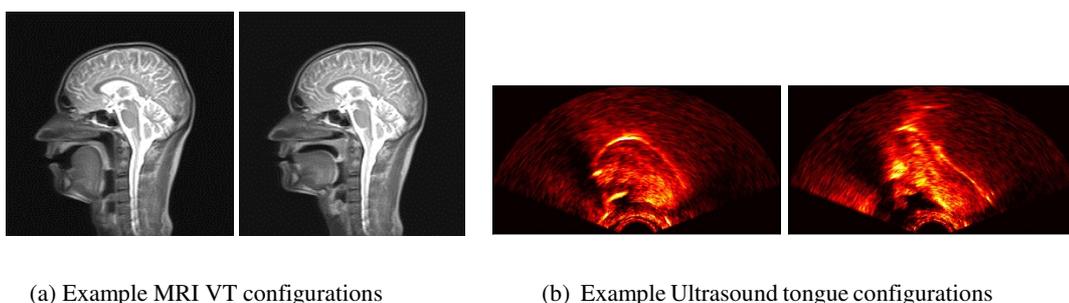

(a) Example MRI VT configurations

(b) Example Ultrasound tongue configurations

Figure 2: *MRI and ultrasound scans of 12 adults (52 speech sounds each) has been collected for data-driven modelling.*

References

[1] M. A. Carreira-Perpinán and Z. Lu, "The Laplacian eigenmaps latent variable model," in *Proc.11th International Conference on Artificial Intelli- gence and Statistics (AISTATS)*, 2007, pp. 59–66.

[2] D. Simon, *Optimal state estimation: Kalman, H [infinity] and nonlinear approaches*. John Wiley and Sons, 2006.

Poster 28:

A single speaker study of Modern Greek:The effect of stress, syllable position and consonantal context on vowels, and of vowel context on intervocalic consonants

James M Scobbie[1] and Anna Sfakianaki[2]

[1]QMU & [2]Aristotle University of Thessaloniki

Introduction

Modern Greek has a typical five-vowel system, /i, e, a, o, u/. Several studies, mainly acoustic, have examined the effect of stress and consonantal context on vowel quality. However, there is no consensus regarding the influence of these parameters on vowel quality, as some of the studies report minimal differences between stressed and unstressed vowels (e.g., Dauer, 1980; Arvaniti, 1994, 2000), whereas others document significant effects on vowel height and backness due to stress (e.g., Fourakis et al., 1999; Nicolaidis & Rispoli, 2005; Baltazani, 2007; Nicolaidis & Sfakianaki, 2007). Vowel sensitivity to consonantal effects has been found minimal along the F1 axis (tongue height) and more substantial along the F2 axis (tongue anteriority) in previous studies of coarticulation (i.e., EPG: Nicolaidis, 1997; acoustic: Sfakianaki, 2012).

This pilot study uses ultrasound tongue imaging (UTI) to investigate these topics from an articulatory perspective. The speaker is the 2nd author. We are aware of just one other UTI study on Greek, which studies the relative timing of consonants in clusters (Yip, 2012).

Protocol

Three blocks of stimuli were recorded from one Greek female adult. Each item was repeated six times, and 216 tokens were analysed in Articulate Assistant Advanced (Articulate Instruments, 2012). Tongue data was recorded at 100fps at QMU with the high-speed Ultrasonix system, using AAA. Synchronised acoustics were captured, with a video image of the lips (de-interlaced to 60fps). The speaker wore a headset (Articulate Instruments 2008). Data has been rotated to the occlusal plane using a bite plate (Lawson et al 2013).

Materials

Single pseudowords with no carrier as follows, to avoid coarticulatory effects with adjacent words or from lingual consonants, and to maximize target segment distinctiveness.

a.   the five vowels /i, e, a, o, u/ sustained in isolation

b.   the five vowels in word-like disyllables of the form /ˈpVpV/ or /pVˈpV/

c.   the three corner vowels /i, a, u/ in disyllables of the form /ˈVCV/, C= /t, s, n, k, x, l, r/.

Results

Stress and position of the vowel in the disyllable (word-initial vs. word-final) do not seem to play an independent or combined effect on the overall tongue shape, which is highly consistent. For example, the first /a/ and the second /a/ in /ˈpapa/ are highly similar in terms of tongue contour. There were however differences between the vowels in the pVpV pseudowords vs. the sustained vowels in isolation. See Figure 1a for average vowel shapes, each mean tongue shape flanked by 1 s.d.
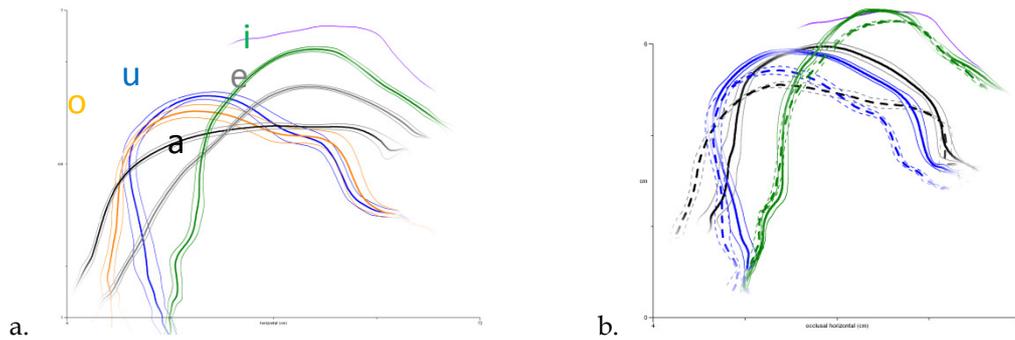
Figure 1 a. vowels (/i/ green, /e/ grey, /a/ black, /o/ orange, /u/ blue) **b.** variants of /k/(solid lines) and /x/ (dashed lines) in 3 colour coded vowel contexts, /i/, /u/ and /a/.
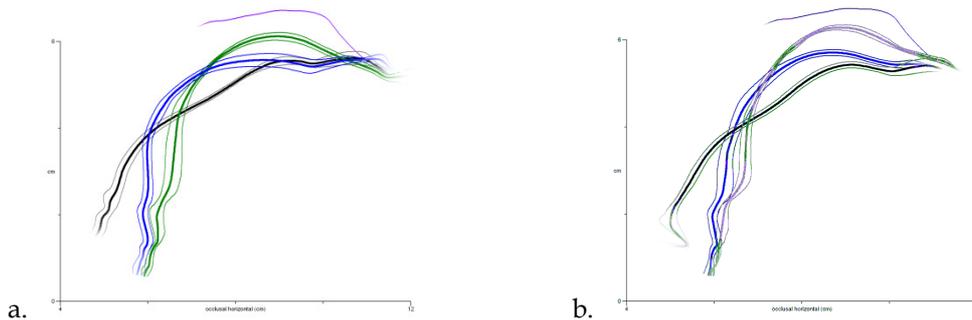

Figure 2 a. Vowel-induced variants of /t/. **b.** vowel-influenced variants of /n/.
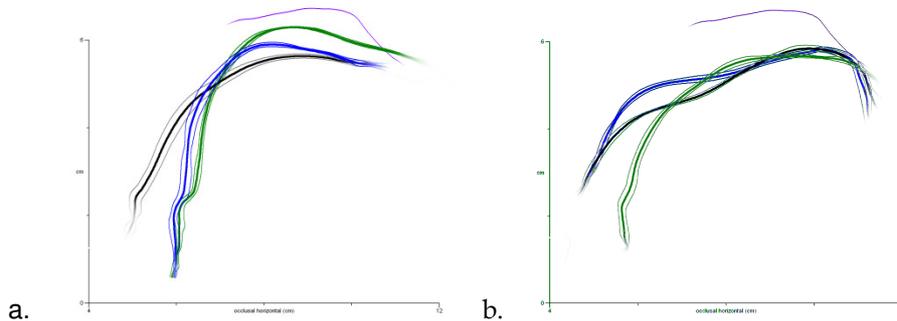

Figure 3 a. Vowel-induced variants of /l/ b. Vowel-induced variants of /s/

In 'VCV, consonantal effects seem to be stronger on the second, weak, vowel than on the first, stressed one, suggesting a higher degree of coarticulation between the consonant and the word-final unstressed vowel. Greater C-to-V effects on unstressed vs. stressed vowels were seen in a previous acoustic study of coarticulation in Greek disyllables, in which higher coarticulatory magnitude was reported in the carryover direction (Sfakianaki, 2012). This finding may reflect the preferred V.CV rather than VC.V syllable pattern for Greek in this context (Joseph & Philippaki-Warburton, 1987).

The consonants in VCV contexts also show some interesting features. Figure 1b shows the the palatal pair of allophones of /k/ and /x/ related by a simple relative constrictional weakening, while the stop/fricative distinction in the back allophones differ greatly in place of articulation: /x/ is uvular in /uxu/ and almost pharyngeal in /axa/. Figures 2a and 2b. show obvious palatalisation of /t/ and /n/ respectively in /i_i/. The also show pharyngealisation in /a_a/. Both consonants pattern with fronted tongue root for /i/ and /u/. The lateral /l/ (Figure 2a) is also like /n/ and /t/. Figure 3b on the other hand shows a tip-down /s/ with different coarticulatory patterns. There is more coarticulatory resistence for /s/ than /t/ and /n/ (as might be expected) in the anterior tongue region (Zharkova, Hewlett and Hardcastle, 2012). While the differences in the tongue root are just as large as the other coronal consonants, surprisingly they are different in organisation: /usu/ patterns with /asa/ this time, in having root retraction, rather than advancement like /isi/.

We do not know the extent to which these articulatory patterns are idiolectal or typical of Modern Greek, nor have we presented the acoustics of this speaker. However, the patterns shown are highly consistent, with low token-to-token variation, and reflect results from previous acoustic studies that show little acoustic variation due to stress or word position.

References

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

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

Arvaniti, A. (1994). Acoustic features of Greek rhythmic structure. *Journal of Phonetics, 22,* 239-268.

Arvaniti, A. (2000). The phonetics of stress in Greek. *Journal of Greek Linguistics, 1,* 9-38.

Baltazani, M. (2007). Prosodic rhythm and the status of vowel reduction in Greek. *Selected Papers on Theoretical and Applied Linguistics from the 17th International Symposium on Theoretical and Applied Linguistics, 1,* 31-43. Department of Theoretical & Applied Linguistics, Aristotle University of Thessaloniki.

Dauer, R. M. (1980). *Stress and rhythm in Modern Greek.* Doctoral Dissertation. University of Edinburgh, Edinburgh, UK.

Fourakis, M., Botinis, A. & Katsaiti, M. (1999). Acoustic characteristics of Greek vowels. *Phonetica* 56, 28-43.

Joseph, B. & Philippaki-Warborton, I. (1987). *Modern Greek.* Kent: Croom Helm Ltd.

Lawson, E. & Scobbie, J.M. (2013) Bite plate: a tool for standardising probe orientation to improve UTI interpretation. Poster presented at Ultrafest6, Edinburgh.

Nicolaidis, K. (1997). *Sources of articulatory variability in Greek: an electropalatographic study.* Doctoral Dissertation. University of Reading, Reading, UK.

Nicolaidis, K. & Rispoli, R. (2005). The effect of noise on speech production: An acoustic study. *Studies in Greek Linguistics, 25,* 415-426.

Nicolaidis, K. & Sfakianaki, A. (2007). An acoustic analysis of vowels produced by Greek speakers with hearing impairment. *Proceedings of the 16th International Congress of Phonetic Sciences,* ID 1358, 1969-1972. Saarbrücken.

Sfakianaki, A. (2012). *An acoustic study of coarticulation in the speech of Greek adults with normal hearing and hearing impairment.* Unpublished Doctoral Thesis. Aristotle University of Thessaloniki.

Yip, J. (2012) Linguistic Effects on the Timing of Gestural Coordination in Modern Greek CC Sequences. *Poster, 164th Meeting of the Acoustical Society of America.* Kansas City, MO: 22–26 October 2012.

Zharkova, N., Hewlett, N. & Hardcastle, W. J. (2012). An ultrasound study of lingual coarticulation in /sV/ syllables produced by adults and typically developing children. *Journal of the International Phonetic Association,* **42,** 193-208.

Poster 30

Assessment of head reference placement methods for optical head-movement correction of ultrasound imaging in speech production

Kevin Roon[1,2], Eric Jackson[1], Hosung Nam[2], Mark Tiede[2], D. H. Whalen[1,2]

[1]CUNY Graduate Center, [2]Haskins Laboratories

One method of quantification of tongue movement using ultrasound imaging during speech production requires determination of tongue position relative to the palate, corrected for probe and head motion so that successive frames can be meaningfully compared. This method involves placing infrared emitting diodes (IREDs) on a 'tiara' attached to the participant's head (Whalen et al., 2005). An alternative is to attach IREDs directly to the participant's skin. In either case, the IREDs can potentially

move relative to the participant's skull. The present study examined movement with both methods for simple utterances, a read paragraph, and spontaneous speech. The amount of IRED movement observed using both methods allowed identification of regions where IREDs should be affixed on a participant's skin to minimize movement when the direct application method is used. Results of simulations showing the effects of this IRED movement on the calculated head-movement correction of the tongue images are presented. Given the results of these simulations, guidelines are proposed for establishing thresholds that can be used to determine whether a given experimental trial should be included based on the amount of reference IRED movement. Differences in movement due to linguistic content or style will also be discussed.

Lingusistic Analysis with Ultrasonix Ultrasound Systems

More than 200 research facilities around the world use Ultrasonix-branded ultrasound systems for a variety of applications including linguistic research

In addition to a range of diagnostic ultrasound systems, Ultrasonix offers a research model, which enables recording of several minutes of very high frame rate images of the tongue. Ideal for use in speech therapy, the transducer is simply positioned under the chin while the patient is speaking. The images are captured and fed into a movement analyzer, providing real-time results to the clinician and patient as they focus on speech training. By visualizing their speech in real-time, the patient is able to learn to correct abnormal speech patterns. Ultrasonix Systems are also compatible with research tools such as AAA from Articulate Instruments.

Menu

8.30 - 9.00am Tea & coffee, still and sparkling water.

10.00-10.30 Morning Break **Wednesday:** Freshly baked Lemon & blueberry muffins **Thursday:** Warmed French pastries **Friday:** Handmade breakfast flapjacks Premium fresh-roasted coffee & speciality and herbal teas orange juice, still and sparkling water

12.30-1.00 Hot lunch

Wednesday:

- Red wine beef Bourguignon with pancetta & mushrooms
- Lamb Moussaka
- Chicken & Parma ham encroute
- Aubergine parmigiano (layers of baked aubergine with mozzarella and tomato ragu)

Thursday:

- Beef lasagna made with layers of rich Bolognaise and Béchamel sauce
- Spiced Tagine of lamb served with Morrocan couscous
- Spiced Thai coconut chicken curry
- Mushroom and artichoke lasagna (V)

Friday:

- Sweet & sour chicken
- Beef & vegetable chilli served with rice or tortillas, sour cream, salsa, cheese & guacamole
- Luxury fish pie topped with buttery mash
- Mushroom in a creamy Madeira sauce wrapped en croute (V)

All of the above served with appropriate vegetable, salad and side accompaniments

* Vegan, gluten free and dairy free options will be available to those who requested them at the time of registration

Soft drinks, orange juice, still and sparkling water

14.30-15.00 (3.00-3.30 on Thursday) Afternoon break

Wednesday:

Chocolate fudge cake

Thursday:

Lemon tart

Friday:

Strawberry & raspberry cheesecake

5.45 - 7.30 Wednesday Reception

Finger food

Rosé and white sparkling wine

Soft drinks, orange juice, apple juice, still and sparkling water

7.00 - late   Friday Dinner & Ceilidh

Venue: Ghillie Dhu, 2 Rutland Place , EH1 2AD See map below

live ceilidh band & lessons

Starters
Roast Tomato & Red Lentil Soup with Crusty Bread
Smooth Chicken Liver Pate in Bacon, Red Onion Marmalade & Arran Oaties
Locally Caught Haggis Neaps & Tatties, Creamy Whisky Gravy

Mains
Baked Chicken Supreme, Peppered Asparagus Farce, Chive Butter Sauce, Seasonal Vegetables &
Creamy Mash
Rolled Fillet of Plaice, Smoked Salmon & Lemon Mousse, Seasonal Vegetables & Creamy Mash
Brie Broccoli, Plum Tomato & Bean Wellington, Seasonal Vegetables & Creamy Mash

Desserts
Traditional Scots Cream Crowdie Cranachan, Toasted Oats Sweet Raspberries & Shortbread
Crumble
Oo'r Tipsy Laird, Vanilla Crème Drambuie Soaked Sponge & Sweet Berry Compote
Sweet Mango Bavarois with Passion Fruit Sauce