Workshop: The use of ultrasound in L2 Ph-research and in teaching non-native speech sounds

Kevin Roon
kdroon@nyu.edu
Phonetics and Experimental Phonology Lab
Department of Linguistics
New York University

CASPSLaP 2012
Friday morning, 17 February 2012, 09:00-10:30
Thanks to

• Eric Holt and the organizers of CASPSSLaP
• Duncan Howe and the USC School of Medicine Ultrasound Institute
• Lisa Davidson (NYU) for slides on ultrasound background and theoretical research summary
• Doug Whalen (CUNY Graduate Center)
• and you for your interest
Workshop Plan

1. Explanation of technology
2. Using ultrasound in research
3. Using ultrasound specifically for teaching of Spanish and Portuguese phonetics and phonology
4. Hands-on use of machines
What problem are we confronting?

• Vowels are described phonologically in terms of height, front/backness, and lip rounding.
• Consonants are described in terms of place and manner of articulation, and voicing.
• Articulatory reality is more complicated.
• It’s really hard to see what the tongue is doing while people are speaking.
Speech production often involves the a lot of the tongue

- Vowel “height” and “frontness” are not straight-forwardly correlated with the location of a single point on the tongue dorsum.
- Secondary articulations of vowels and consonants in English may not be appropriate in another language.
  - Advanced Tongue Root in English /i/
  - Tongue-back gesture for English /l/
- Sounds in other languages can require us to do things we don’t do in English.
  - Trilled /r/
  - Dental coronals /t̪, d̪, ̃n/ (some New Yorkers excepted...)
Ultrasound imaging

• Dynamic tongue movement can be reconstructed by collecting images of soft tissue (the tongue) from the echo patterns of ultra-high frequency sound.

from Maureen Stone, Vocal Tract Visualization Lab, UMB
http://speech.umaryland.edu
How does it work?

• Ultrasound transducer emits ultra-high frequency sound waves that travel through soft tissue.

• Abrupt changes in tissue density cause echoes of these sound waves.

• Ultrasound machine uses these echoes to construct images of objects.
How it works (continued)

• The transition from the tongue to air is a very abrupt change in density.

• Surface of the tongue therefore images well with ultrasound.
Video

• Placing an ultrasound transducer under the jaw can provide a real-time video of the movement of the surface of the tongue (mid-sagittal view, most typically).

• Video can be seen on the machine’s built-in screen, projected, and/or recorded to hard-drive.

• A regular machine can capture about 33 frames per second.

• A really good (expensive) machine can get ~100 fps.
Ultrasound

AS A RESEARCH TOOL
Advantages of ultrasound

- Unique in its ability to provide a 2D image of most of the tongue surface.
- Decent time resolution frame-capture rate
- Non-invasive
- Non-toxic
- Relatively cheap
- Generally accessible if not acquirable
- Portable
- Record direct to computer for subsequent analysis.
Qualitative information and fieldwork

• Articulation of clicks in Khoisan languages (Miller, OSU)

• Shapes of English /r/ (Campbell, UBC; Scobie, QMUC; Baker & Mielke, U. Arizona)

slide credit: Lisa Davidson
Quantitative phonetics research

- 3D tongue shapes (Stone, University of Maryland, Baltimore)
Quantitative sociophonetics research

- [æ]-tensing in a New Jersey dialect (De Decker and Nycz, NYU)
- What is the acoustic tensing in [pæn] (as compared to [pæs] or [pæd]) due to?
  - Nasalization hypothesis, articulatory hypothesis, (or both?)
Quantitative phonology research

• Repair of non-native consonant clusters (Davidson 2005)
  – Is the schwa in an English speaker’s repair of *zbura* as *[zəbura]* the same as a lexical schwa?
  – Compare words like *[zə]bura* to “spurt” & “superfluous”
  – Results: the schwa in *[zəbura]* is more similar to the production of “spurt” than “superfluous”

*slide credit: Lisa Davidson*
What ultrasound is not good for

• Fine time resolution
  – Ultrasound has a frame rate of 33fps, which isn’t much considering that some sounds are less than 30ms long! Though some machines now have closer to 100fps.

• Point tracking
  – Ultrasound does not track a single point over time

• Other vocal tract structures
  – To see pharynx and larynx

• Constrictions: air above tongue means we can’t see the palate

• Nasals: or the velum

image credit: Mark Tiede, Haskins Labs

slide credit: Lisa Davidson
Some other limitations

- Ultrasound cannot image when it encounters **bone** or **air**.
- The hyoid and jaw bones cast posterior and anterior shadows that usually delimit the imageable region of the tongue.
- The tongue tip often has a pocket of air under it, which means it can be hard to image.
- Image can be fuzzy, depends on the individual.

*image credit: Diana Archangeli and Jeff Mielke, University of Arizona*

*slide credit: Lisa Davidson*
Technical issues for research

• Head and transducer stabilization
  – If the head and transducer move around, then we cannot compare tongue shapes
  • E.g., is /i/ higher than /e/? What if the person raises his head when producing /e/?
  • Need same relative space
Head and transducer stabilization (NYU)

- Moldable head stabilizer is attached to the wall with Velcro.
- Head is further stabilized with Velcro.
- Transducer held rigid with Magic Arm

Kevin Roon, CASPSLaP 2012
Technical issues for research

The images are interesting, but we need to get numbers out of them.

• “Edge tracking”: quantifying the surface shape.
• Statistical analyses: depends on what is being studied and where differences lie.
• Relatively easy to show that two shapes are different (e.g., smoothing spline ANOVA, Davidson 2006)
• Less easy (but possible) to quantify how they are different.
Growing community of researchers

- Growing number of researchers and institutions using ultrasound in speech research: U Maryland, Haskins Labs, UBC, Queen Margaret University (Scotland), CNRS Grenoble, NYU, U Arizona, USC, U South Florida, UQAM...and growing.

- Ultrafest: regular meeting of researchers who use ultrasound in the investigation of speech
Ultrasound Applied

FOR TEACHING SPANISH AND PORTUGUESE
The basic proposition: visual feedback

• When trying to produce speech sounds that are new to them, learners normally get the following feedback:
  – acoustic input, from themselves and native speakers
  – limited visual input about articulation, mostly from speakers they observe
  – proprioceptive feedback from themselves, but
    • for many people it is hard to introspect on
    • useful only if they know what they’re supposed to be doing in the first place.

• Ultrasound can provide real-time, visual feedback of the tongue, complementing all of the above.

• This should be helpful.
Advantages of ultrasound

- Real-time video feedback
- Allows teacher and subject to see what the tongue should (and shouldn’t) be doing
- Reasonably intuitive to interpret the image
- Non-invasive
- Non-toxic
- Relatively cheap
- Generally accessible if not acquirable
- Portable

Kevin Roon, CASPSLaP 2012
Problem type #1

• Articulation of speech sounds that simply do not exist in the learner’s native language:
  – production of the Spanish fricative /x/
  – Spanish trilled /r/
  – the high, back, unrounded Iberian Portuguese vowel /ɯ/
  – dental (vs. alveolar) Spanish stops /d, t, n/
  – any others?
Spanish /x/

With Ultrasound:

• Student produces /h/ (assuming this is the mistake they make).

• Instructor produces /x/: this should look markedly different immediately.

• Student produces /k/: that this is closer to /x/ than /h/ should be obvious

• Student works on /x/, with instructor demonstration as/if needed.
The Holy Grail: trilled /r/

Bad news:

• The tongue tip makes contact with the alveolar ridge in the trill, and ultrasound isn’t very good at imaging the tongue tip.
• The tongue tip moves too fast for the ultrasound to see given its sampling rate.
• The tongue tip is actually a passive articulator anyway, so seeing what it does probably won’t help a learner produce it.
The Holy Grail: trilled /r/

All is not lost:

• Spanish /r/ has many components.
• Iskarous et al. The tongue back has a bracing gesture that allows the tongue tip to trill. This gesture is more readily visible.
• The tongue tip is actually a passive articulator anyway, so seeing what it does probably won’t help a learner produce it.
Problem type #2

- Getting rid of articulatory “baggage” that the learner’s native phonology brings to the language being learned:
  - producing vowels like /e/ without the off-glide that is obligatory in English
  - producing Portuguese nasal vowels without any following coronal “n”-gesture
  - producing a velarized (“dark”) l [ɬ]) word-finally
Spanish /e/, English [eɪ]

• /e/ exists in English only as a diphthong, that is, it always has the off-glide [eɪ].
• In Spanish it’s a monophthong [e].
• English speakers have enormous difficulty producing that vowel without the off-glide.
• They tend to be very deaf to the difference, that is, to them, [eɪ] sounds the same as [e].
• How can they learn the right articulation if they can’t even hear the difference?!?
Spanish /e/, English [e\textsuperscript{I}]

Setup before Ultrasound:

- Instructor produces “José” [xose], asks student to do same.
- Student (most likely) produces [ho\textsuperscript{0}ze\textsuperscript{I}]. (Note a similar off-glide problem exists for /o/.)
- Attempts to point out the difference in the last vowel based on acoustics only may get quizzical looks from student.

With Ultrasound:

- Instructor produces [xose]
- Student prepares to say the word using ultrasound, but is instructed to watch for the quick movement of the tongue up and forward at the end of the word.
- Student thinks instructor is crazy.
- Student will, hopefully, be able to see what he/she cannot hear, and use this feedback to get rid of the off-glide.

Kevin Roon, CASPSLaP 2012
Practical idea: #1

Interactive, real-time instruction

– Instructor and student(s) use ultrasound together.
– Instructor demonstrates, student attempts to replicate.
– Ideally more than one machine at the same time, but possible with just one.

Pros: Interactive  
Cons: Labor intensive

Individualized  
Hard to refer back later

Kevin Roon, CASPSLaP 2012
Practical Idea #2

“Language lab” approach
   – Create recordings of inventories of sounds as exercises that have temporally integrated audio, facial video, and ultrasound video.
   – Frisch et al. at U Southern Florida for English have done this for English (for working with patients with speech pathologies).
   – With an ultrasound machine and a mirror, learners could practice.

Pro: Repeatable
Cons: Does not exist (yet!)
Teacher needn’t be present
Needs dedicated machine

Big opportunity...?
Thanks for your attention!

Time now for interactive ultrasound

HANDS-ON USE OF THE MACHINES
A few useful links

• Ultrasound research at the Vocal Tract Visualization Laboratory of the University of Maryland: http://speech.umaryland.edu/research.html#ultrasound


• The University of Arizona Phonological Imaging Lab (has some good links, too): http://apil.arizona.edu/

• The last ultrafest, held at Haskins Laboratories: http://www.haskins.yale.edu/conferences/UltrafestV/