Ultrasound Imaging in Speech Research

Pertti Palo

11 Nov 2020

Outline for today

- First half: Basic properties and background
 - Advantages
 - Limitations
 - How does it work
 - Anatomy of an ultrasound tongue image
- Second half: Data analysis what can we get out of it?
 - First peek at actual data
 - Videos
 - Splines
- Possible extras
 - Computational measures
 - How a session is run

First half:

Basic properties and background

Advantages

- ► Very safe no known health risks when operated at low power.
- Easy to use after some initial setup.
- Undemanding on the participant as long as the sessions are not long.
- Relatively cheap.
- Frame rate on regular systems 20-120 Hz (and up to almost 400 Hz on exceptional systems), i.e. relatively good time resolution

Limitations I: Method

Limited image quality

- Time resolution is inversely related to spatial resolution.
- Resolution also depends on probe frequency being used: Low frequencies penetrate well, but provide poorer resolution, high frequencies give better resolution but do not penetrate so deep in tissues.
- Especially to unaccustomed eyes the pictures are far from clear.
- Limited range of structures that can be observed.
 - ► Generally nothing behind a bone or an air gap can be imaged.
 - Complex tissue internal features may also make it difficult to meaningfully interpret images.

Limitations II: People

- Choice of participants:
 - Very small children can not use the helmet and tend to move a lot.
 - ► Hand held data is possible but more difficult to analyse.
 - Big headed people are generally difficult to image.
 - A lot of fatty tissue under the chin makes imaging more difficult.
 - Age adds connective tissue inside the tongue and blurs the image.

How does it work?



- ► The probe is a crucial component: Sender-receiver unit which utilises an array of piezoelectric crystals.
- Emits sound pressure waves of very high frequency (2-12 MHz, 1MHz = 1000kHz = 1 million Hz)
- The waves travel through soft tissue and reflect back when they reach a border to structures with different density (bone, air ...)

Anatomy of an ultrasound tongue image I

A regular – fairly clear – ultrasound image of the tongue



Anatomy of an ultrasound tongue image II

Side path: MR image - not an ultrasound image



Anatomy of an ultrasound tongue image III

Down the side path: Vocal tract outline extracted from MRI with guestimated teeth.



Anatomy of an ultrasound tongue image IV

A rotated ultrasound image matched to the MRI outline



Anatomy of an ultrasound tongue image V

And finally the ultrasound image in the 'correct' orientation.



Second half:

Data analysis - what can we get out of it?

First peek at actual data

I'll fire up AAA in a moment and we'll have a look. But let's first talk about what we are going to look at.

What could we have a look at?

- ► Tommi Nieminen & Pertti Palo: Suomen švaan artikulaatio.
 - ▶ Read Finnish sentences with and without epenthetic vowels.
 - Small pilot: one unreliable speaker (me).
 - Good time and space resolution, clear images.
- Pertti Palo & Sonja Dahlgren: Pilot data on Finnish coarticulation type.
 - Read Finnish words and non-words.
 - Publishable pilot: Two reliable speakers.
 - ► Good time and space resolution, reasonably clear images.
- Jalal Al-Tamimi & Pertti Palo (forthcoming): "Tongue contours in guttural consonants in Levantine Arabic: A Generalised Additive Modelling Approach"
 - Read Levantine Arabic words.
 - Large data set which will be the basis of at least one journal article: 10 speakers.
 - Good time resolution, average spatial resolution, not the clearest images.

Just the videos

- > Ultrasound videos can be annotated like any other videos.
- Use either AAA or export the videos and use a video annotation program.
- ▶ It is also possible to do direct measurements on the images.

Splines I



- Splining means we draw a contour on the image and extract it.
- ► Can be done either manually or automatically.
- ▶ Results depend a lot on the quality of the data.

Splines: the results



- Be traditional get point measures in time from two or more contexts and compare them.
- Be modern get movement sequences splined and analyse the data with functional data analysis.
- Or perhaps run shape metrics on single frames or sequences of splines.

That's it!

Thank you!

And thanks to Felix Schaeffler for some of the images and text in these slides, and to Steve Cowen for the picture of me.

References

Al-Tamimi, J. and Palo, P. (2020).

Tongue contours in guttural consonants in Levantine Arabic: A Generalised Additive Modelling approach (provisional title).

In preparation.

Dahlgren, S. and Palo, P. (2021).

Studying language-specific coarticulatory patterns with tongue ultrasound: The case of Finnish.

In preparation.

Computational measures



- We can also analyse the information content of the videos without extracting features such as the tongue contour. For example:
 - Pixel difference tracks over all change and is based on using raw ultrasound data.
 - Optic flow tracks movement by statistically matching local features in the videos.
- These methods do not replace splining, but rather provide options.

Raw data



How a session is run

- Clean everything. This is medical equipment, people are dirty and the plague kills.
- Inform your participants, do the paper work, show them how it works.
- Fit the helmet.
- Run data.
- Tell your participant they've been a great help and help them get out of the helmet.
- ► Also help them get the blue stuff off themselves.
- Clean everything.