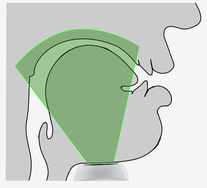
A method for distinguishing tongue surface topology for different categories of speech sound

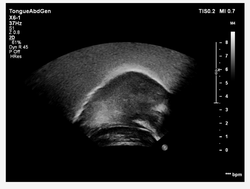
Jonathan Washington, *Swarthmore College* Paul Washington, *Marietta College*

https://github.com/jonorthwash/ultrasound-processing

The problem

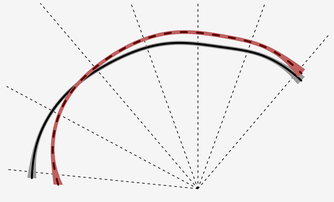
How to compare ultrasound tongue imaging data between individuals and languages?

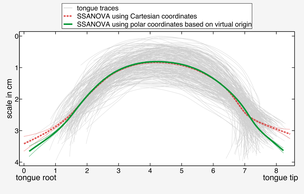




- Ultrasound images of the tongue have no fixed reference point
- Challenging to compare categories (e.g., front/back vowels)
- Quantitative comparison of speakers or languages requires a reference system
- The proposed method yields a dynamic reference system that makes comparison of speakers and languages possible

Existing approach: SSANOVA





Smoothing-spline analysis of variances:

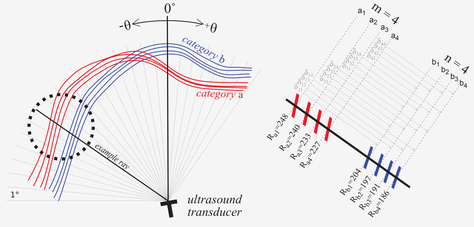
- Originally proposed by Davidson (2006) for tongue trace analysis
- Use of polar coordinates - proposed by Mielke (2015)
- Use transducer as origin - proposed by Heyne and Derrick (2015)

Shortcomings:

- Misleadingly "tight" confidence intervals
- Used for comparing two categories for a single speaker (doesn't expand well to comparisons btw. speakers or languages)
- Normally assessed visually, no quantification

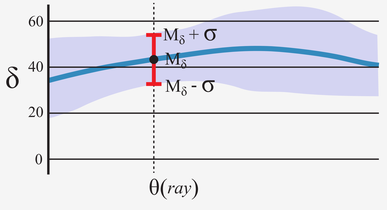
Comparing distinctness of tongue surface topologies

$$M_\delta[ray\ u] = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(R_{ai} - R_{bj})}{m * n}$$



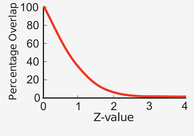
Step 1: Calculate the mean of the radial distances between category a and category b traces

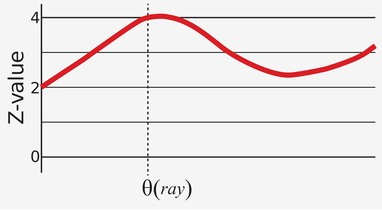
$$\sigma[ray\ u] = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}((R_{ai} - R_{bj}) - M_\delta)^2}{(m * n) - 1}}$$



Step 2: Plot the mean and std dev band (y) vs ray angle (x)

$$Z[ray\ u] = \frac{M_\delta[ray\ u]}{\sigma[rayu]}$$

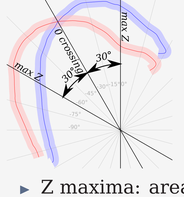




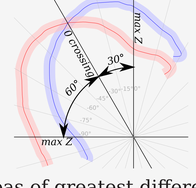
Step 3: Plot Z-value vs ray angle to find areas of most significant difference

Comparing across speakers and languages

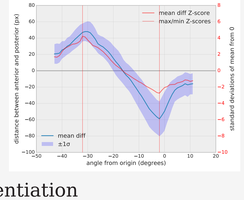
Speaker A: 1:1 ratio



Speaker B: 2:1 ratio



Non-hypothetical speaker

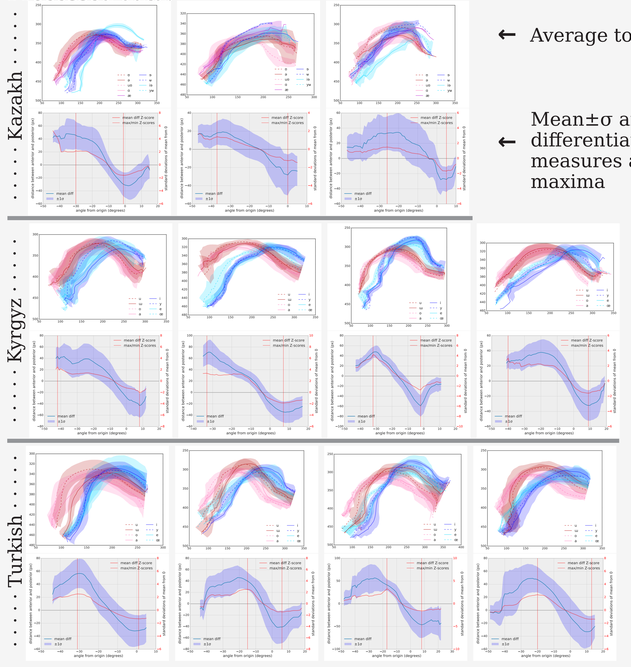


- Z maxima: areas of greatest differentiation
- Z=0 (zero-crossing): area of no differentiation
- **Dynamic reference system**: angular difference between Z=0 and Z maxima as a ratio
- **Easily interpretable, computationally simple**

Application to vowel anterioirity in Turkic languages

Question: Is vowel anterioirity distinguished by the position of the tongue body (TB) or the tongue root (TR)?

Processed data:



← Average tongue traces

← Mean±σ areas of differentiation measures and Z maxima

Analysis (ratios of maxima):

Turkish speakers:

- Ratio of maxima ~1:1
- TB and TR contribute equally to category difference

Kyrgyz and Kazakh speakers:

- Ratio of maxima ≥2:1
- TR contributes more than TB to category difference

Conclusions:

- Turkish uses primarily TB position to differentiate vowel anterioirity (VA)
- Kazakh and Kyrgyz, in addition, actively use TR position to differentiate VA

Acknowledgements

Participants: Madina Satybaldinova, Kishimjan Eshenkulova, Aqsana Borbiyeva, Aybike Şeyma Tezel, Mirlanbek Nurmatov, Saulet Alpysbayeva, Aitbubu Abdyibraeva, Aigul Sakieva, and others who preferred to remain anonymous

Funding: IU University Graduate School Grant-in-Aid of Doctoral Research

Facilities: Indiana University Speech Production Laboratory

References

Davidson, Lisa (2006). "Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance". In: *Journal of the Acoustic Society of America* 120, pp. 407–415. DOI: 10.1121/1.2208133.

Heyne, Matthias and Donald Derrick (2015). "Using a radial ultrasound probe's virtual origin to compute midsagittal smoothing splines in polar coordinates". In: *Journal of the Acoustical Society of America* 138.6. DOI: 10.1121/1.4937168.

Mielke, Jeff (2015). "An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons". In: *The Journal of the Acoustical Society of America* 137.5, pp. 2858–2869. DOI: 10.1121/1.4919346.

SWARTHMORE

INDIANA UNIVERSITY
THE UNIVERSITY GRADUATE SCHOOL

Speech Production Lab