Tongue part movement variability in /r/-final syllables measured by automatically tracking ultrasound images

Sarah R. Li,[1] *Sarah Dugan,*[2,3] *Colin Annand,*[2] *Kathryn Eary,*[1] *Michael Swearengen,*[1] *Gregory Terrell,*[1] *Suzanne Boyce,*[3] *Michael A. Riley,*[2] *T. Douglas Mast*[1]

[1]Biomedical Engineering, University of Cincinnati, Cincinnati, OH, United States
[2]Psychology, University of Cincinnati, Cincinnati, OH, United States
[3]Communication Sciences and Disorders, University of Cincinnati, Cincinnati, OH, United States

lisr@mail.uc.edu

Background: In standard ultrasound biofeedback therapy, midsagittal ultrasound images of the tongue surface are used as feedback on tongue movement (Preston et al., 2017). However, speech outcomes can be mixed (Sugden et al., 2019), potentially due to the complexity of interpreting these ultrasound images. A simplified representation of the complex tongue surface curve might enhance motor learning in therapy by encouraging external focus on target movement patterns (Maas et al., 2008). For simplification, it is useful to quantify tongue movement via tongue surface tracking. For therapy, it is important to compare performance to target patterns. In particular, it is important to identify patterns of movement and characterize typical amounts of variability within speakers and phonetic contexts.

Our focus is the most common speech production error in children: American English /r/. It is well known that different vowels are produced with different tongue movement variability. Characterizing tongue movement variability of /r/ in different vowel contexts, as well as variability for vowels, is therefore potentially significant in determining targets for speech therapy.

We have developed a tongue tracking method, TonguePART, that is computationally efficient and can run in real time with minimal user input. TonguePART tracks movement of the blade, dorsum, and root parts of the tongue independently. Trajectories of movement measured by TonguePART were used to compare articulatory variability for /r/-final syllables of different vowel contexts in child speakers with typically developing (TD) speech or residual speech sound disorder (RSSD).

Methods: TonguePART applies a low-pass filter to the B-mode image and maps the tongue surface as local brightness maxima within search regions estimated by Taylor series expansions (Dugan et al., 2019). For each speaker, a researcher selected a spatial calibration point and two brightness threshold values to track all productions; new calibration and threshold values were changed by vowel context or production if necessary.

Participants were 16 (8 TD and 8 RSSD) children aged 8-17, speakers of a rhotic American English dialect. The stimuli were 4-10 productions each of /ir/, /per/, /ɑr/, /or/, /buər/, and /pɛr/, recorded with a Siemens Acuson X300 PE ultrasound system at 36 fps. Trajectories were temporally interpolated to match durations. Productions were excluded (<20%) if tracking errors were indicated by detected discontinuities or visual inspection.

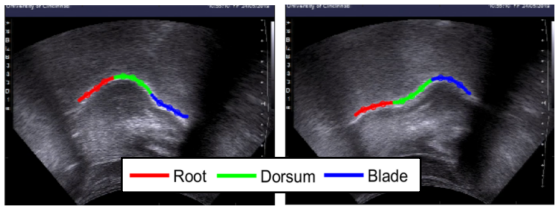

Figure 1: *Tracked B-mode images for the start (left) and end (right) of a selected /or/ production*

We quantify variability of these tongue part displacement trajectories using the time-dependent standard deviation of displacement for each tongue part, as well as the root-mean-square value of this standard deviation across tongue parts, productions, and speakers.

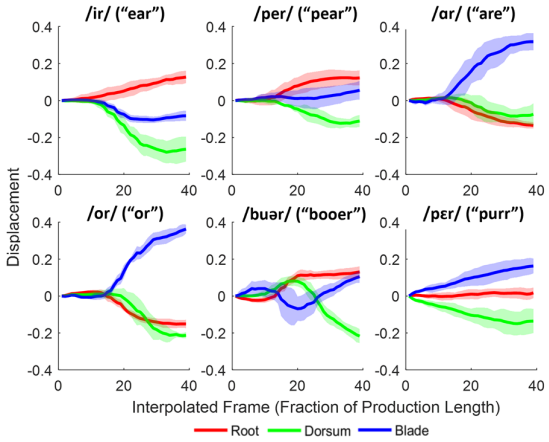

Figure 2: *Trajectories with time-dependent standard deviations (shading) for each /r/-final vowel context of a selected speaker*

Results and discussion: Preliminary results show TonguePART can capture complex, rapid movements of tongue parts; Figure 1 demonstrates an example of tongue tracking for a selected /or/ production. The current MATLAB implementation computes and displays displacements at ~32 fps, indicating real-time capability.

Figure 2 shows representative mean trajectories and intraspeaker variability for one TD speaker. Some TD speakers had greater variability for /ir/ and less for /pɛr/. Compared to TD speakers, RSSD speakers were more inconsistent across vowels, with some vowel patterns in the same speaker showing less variability due to smaller overall tongue movement.

Observed movement variability trends provide insight into articulatory variation. Implications for selection of tongue movement targets for ultrasound-based biofeedback in speech therapy are also discussed.

Acknowledgements: This work was funded by NIH/NIDCD grant 1R01DC017301-01.

References

Dugan, S., Li, S. R., Masterson, J., Woeste, H., Mahalingam, N., Spencer, C., Mast, T. D., Riley, M. A., & Boyce, S. E. (2019). Tongue Part Movement Trajectories for /r/ Using Ultrasound. *Perspectives of the ASHA Special Interest Groups*, *4*(6), 1644–1652.

Maas, E., Robin, D. A., Austermann Hula, S. N., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of Motor Learning in Treatment of Motor Speech Disorders. *American Journal of Speech-Language Pathology*, *17*(3), 277–298.

Preston, J. L., McAllister Byun, T., Boyce, S. E., Hamilton, S., Tiede, M., Phillips, E., Rivera-Campos, A., & Whalen, D. H. (2017). Ultrasound Images of the Tongue: A Tutorial for Assessment and Remediation of Speech Sound Errors. *Journal of Visualized Experiments*, *119*, 55123.

Sugden, E., Lloyd, S., Lam, J., & Cleland, J. (2019). Systematic review of ultrasound visual biofeedback in intervention for speech sound disorders. *International Journal of Language & Communication Disorders*, *54*(5), 705–728.

Keywords: Speech production, Speech motor control, Tongue motion tracking, Rhotics, Speech sound disorders, Visual biofeedback