THE RELATIONSHIP BETWEEN HIGHER FORMANTS AND TONGUE SHAPES IN AMERICAN ENGLISH RHOTICS

Amanda Eads[1]; Heather Kabakoff[2]; Hannah King[3]; Elaine Hitchcock[4]; Jonathan L. Preston[5]; Tara McAllister[1]

PURPOSE

- Because Speech Sound Disorder (SSD) can lead to lasting academic and social challenges [1-2] and effective therapy is difficult to achieve given Speech Language Pathologists' (SLP) large caseloads[3-4], AI tools to augment SLP-delivered services could improve outcomes.
- Residual distortions of /ɹ/ are one of the most common challenges for children with SSD, in part due to its articulatory complexity[5].
- Acoustically, the various tongue shape types for American English /ɹ/ are perceptually indistinguishable[6], but previous work has found differences among higher formant frequencies with lower values corresponding with retroflex shapes[7-9].
- This project aims to create an acoustic classifier that recognizes multiple tongue shape types, with the goal of providing qualitative articulatory feedback to inform clinical cueing strategies.

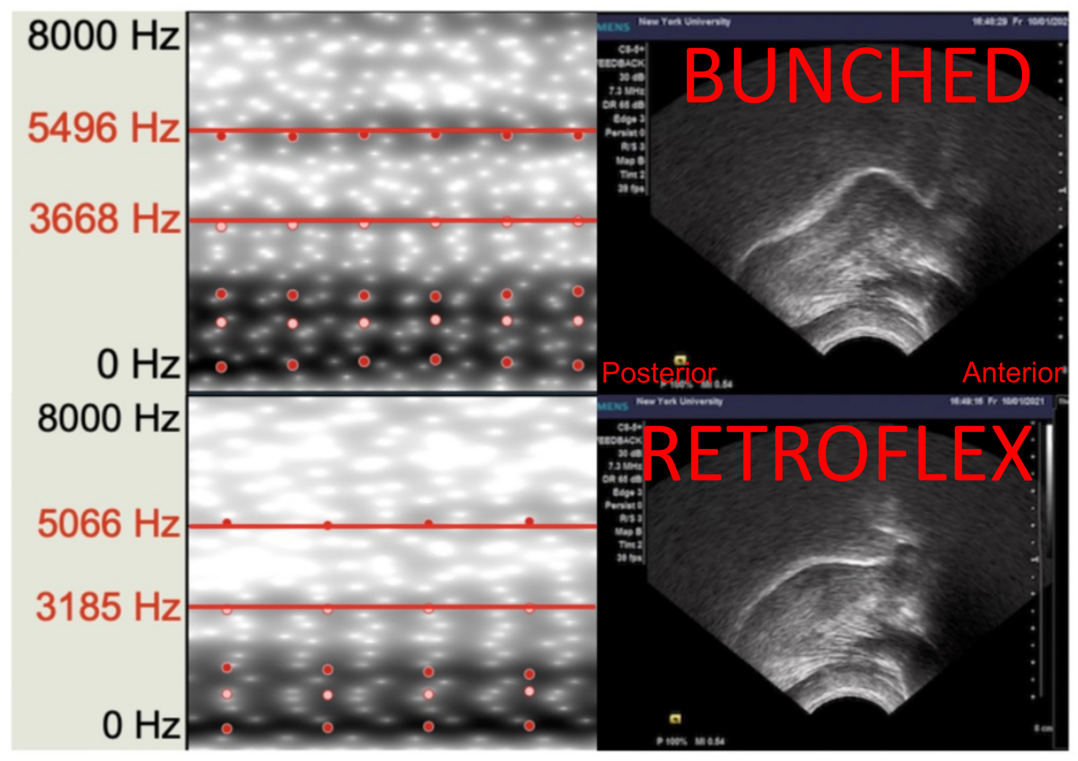
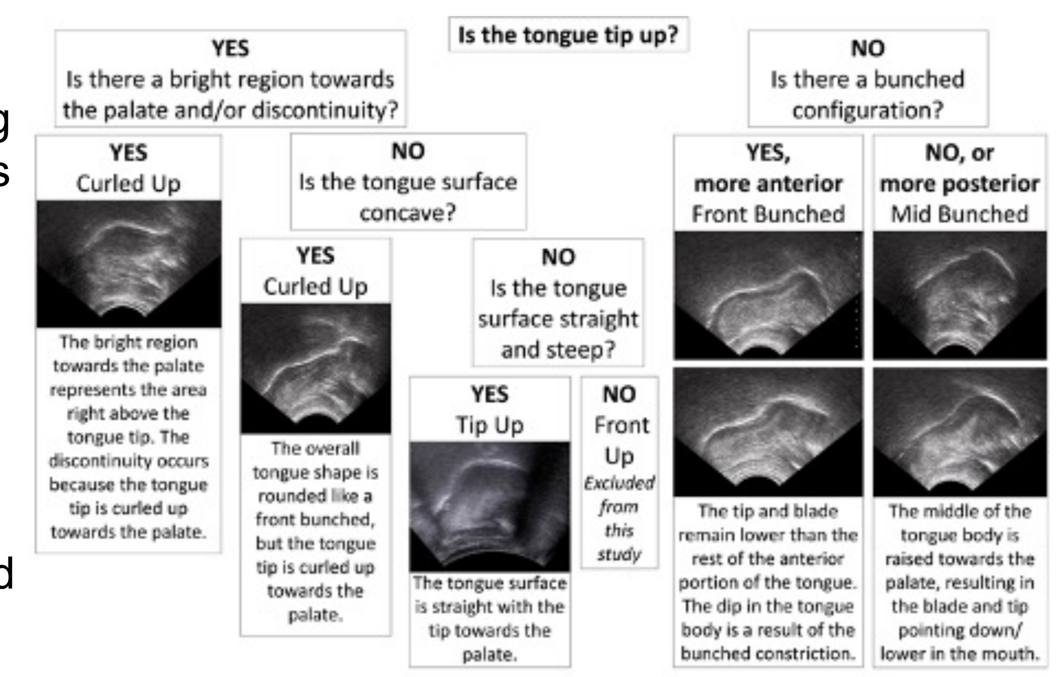


Fig 3: Spectrogram and ultrasound images from one child show acoustic differences (F4 and F5) between retroflex and bunched /ɹ/.

RHOTIC TONGUE SHAPES

- Previous research using the Wisconsin X-ray Microbeam and the University of Cincinnati databases demonstrates 20+ tongue shapes that produce perceptually accurate American English /ɹ/[10-11].
- Shape classification flowchart following King & Ferragne[12] classifies five tongue shape types specifically from ultrasound data:
  - Curled Up
  - Tip Up
  - *[Front Up]*
  - Front Bunched
  - Mid Bunched

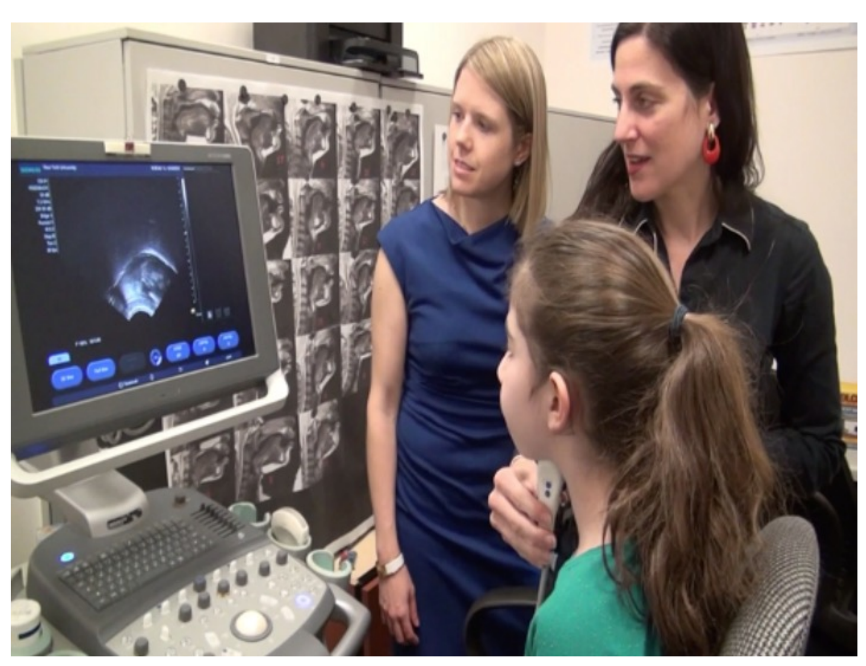


RESEARCH QUESTION

Can we predict rhotic tongue shape type from the acoustics of /ɹ/ in a sample of children aged 9-15 years old?

PARTICIPANTS

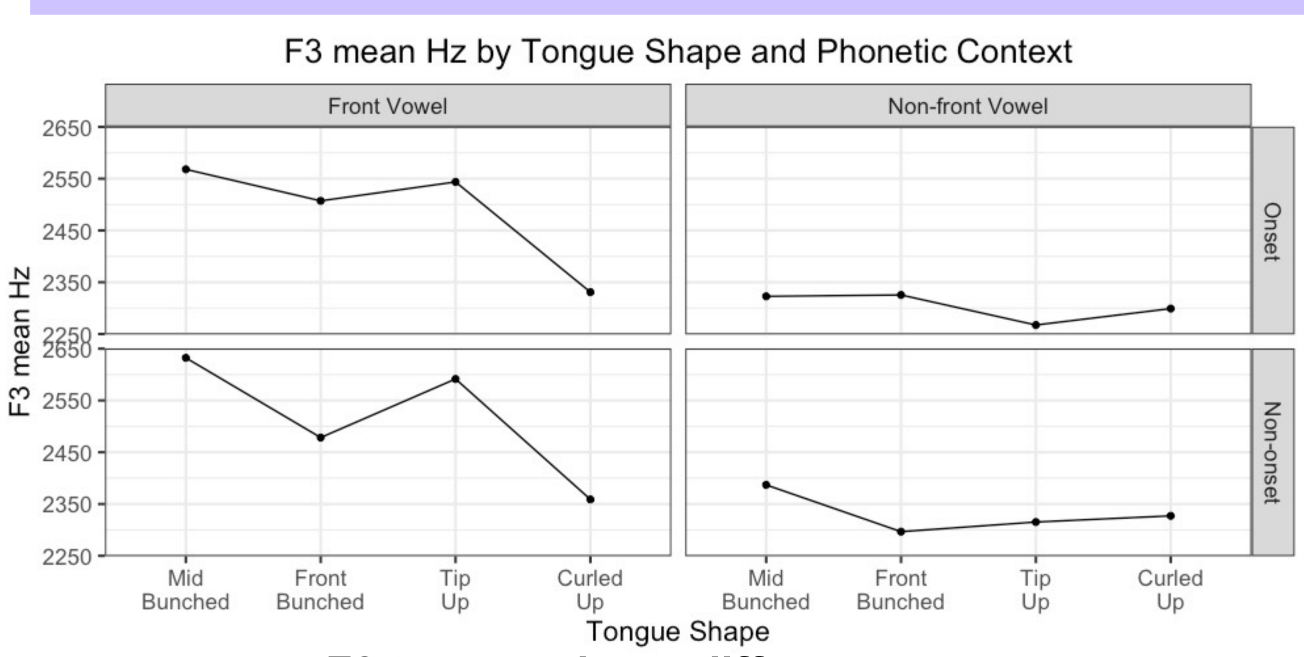
- 30 children (17 male, 13 female) with residual speech sound disorder (RSSD) for /ɹ/ who learned to produce perceptually accurate /ɹ/ during 10 weeks of ultrasound biofeedback treatment (2-3 sessions/wk) at New York University or Haskins Laboratories.
- 36 children (17 male, 19 female) without residual speech sound disorder for /ɹ/.
- Groups were matched in age (9-15 years), spoke a rhotic American English dialect, and had no speech-language-hearing differences other than RSSD.



METHODS

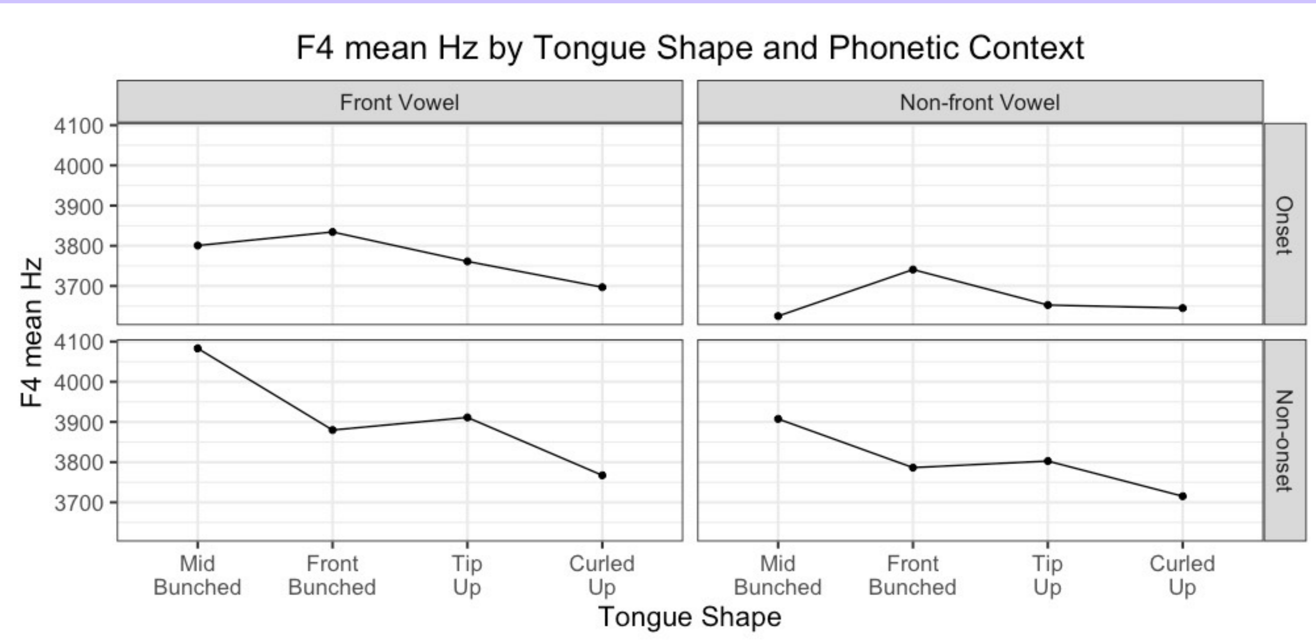
- **Production task:** *Stimulability* probe phonetically balanced by consonantal, syllabic, vocalic /ɹ/
- **Perceptual accuracy ratings:** Masked clinician ratings
- **Tongue shape category:**
  - Ultrasound video (Siemens C8-5 transducer) via video capture card
  - Labeled /ɹ/ interval in Praat[13]; viewed frames in *GetContours*[14-15]
  - Coded tongue shape category using flowchart above[12]
- **Acoustic Measurement:**
  - Montreal Forced Aligner[16] and version 3.0 of the PERCEPT pre-trained acoustic model[17]
  - Adapted the Voweltine script[18] to select best formant settings per token/speaker and measure F1-F5
  - Extracted mean Hertz value from the steadiest 25 milliseconds

RESULTS



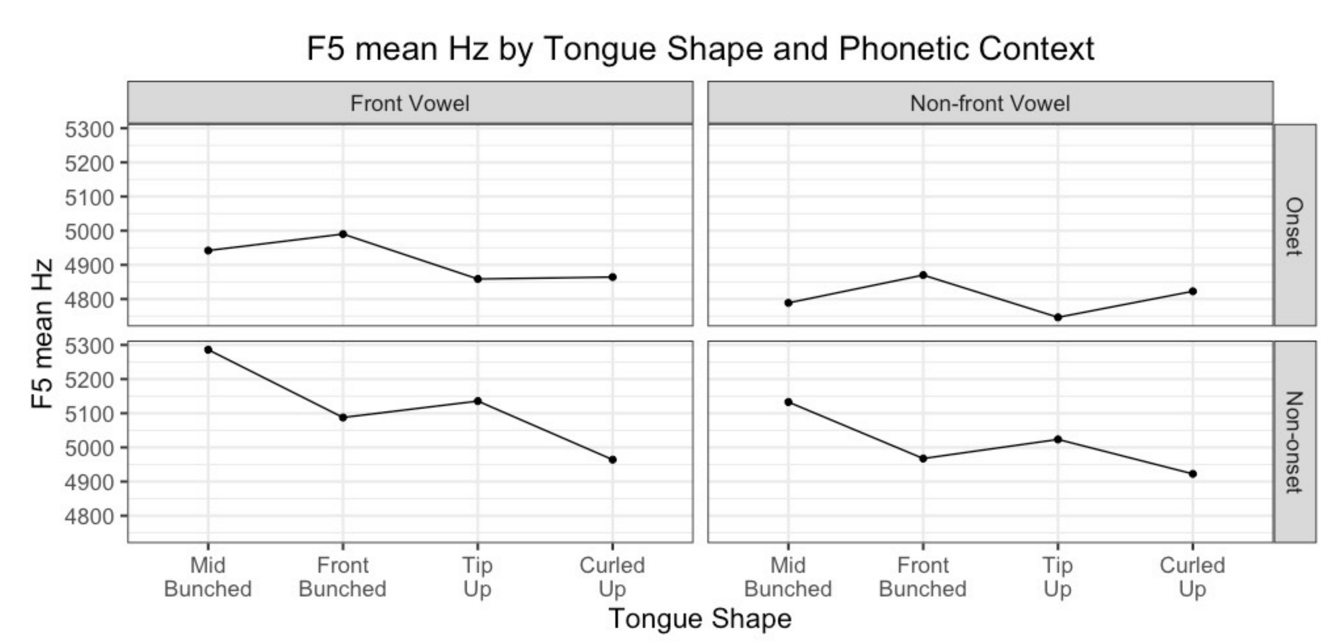
F3 tongue shape differences are more prominent in front vowel contexts.

- Significant effect of vowel context ($p < 0.001$);
- **Curled Up** is significantly different from **Mid Bunched** at F3 mean Hz ($p=0.03$) and trends toward a difference from **Front Bunched** ($p = 0.05$);
- Significant interactions between tongue shapes (**Curled Up**: $p < 0.001$; **Mid Bunched**: $p < 0.01$; **Front Bunched**: $p = 0.03$) and vowel context.



F4 tongue shape differences are impacted by vowel context and syllable position.

- Significant effect of vowel context ($p< 0.01$);
- Significant effect of syllable position: onset versus non-onset ($p< 0.01$);
- Significant interaction between syllable position and the **Mid Bunched and Front Bunched** types ($p < 0.02$) and a trend towards significance between **Mid Bunched and Curled Up** types ($p = 0.05$) indicates differences depend on syllable position.



F5 tongue shape differences depend on syllable position.

- Significant effect of syllable position: onset versus non-onset ($p< 0.01$);
- Trend towards significance for vowel context ($p= 0.07$)
- Significant interaction between syllable position and the **Mid Bunched and Front Bunched** types ($p = 0.04$) and a trend towards significance between **Mid Bunched and Curled Up** types ($p = 0.07$) indicates differences depend on syllable position.

TAKEAWAYS

- Acoustic differences between tongue shapes must be considered in relation to syllable position and vowel context!
- Wondering what's going on with the classic binary bunched-retroflex differences…?
  - Similar findings/interactions
  - Consistent with previous work[7-9], retroflex tongue shapes show lower F4 mean Hz on average
- Many future directions:
  - Clustering analysis on the current data including F3-F5
  - Vocal tract modeling collaboration
  - Investigating acoustics and articulation of perceptually inaccurate /ɹ/
  - Developing the acoustic classifier
  - Releasing the corpus

REFERENCES

1. J. E. Bernthal, N. W. Bankson, and P. Flipsen Jr., Articulation and Phonological Disorders: Speech Sound Disorders in Children, 8th ed. Pearson Education, 2017.
2. J. C. Vick et al., "Data-driven subclassification of speech sound disorders in preschool children," Journal of Speech, Language, and Hearing Research, vol. 57, no. 6, pp. 2033–2050, 2014.
3. R. Koegel, L. Koegel, J. Ingham, and K. Van Voy, "Within-clinic versus outside-of-clinic self-monitoring of articulation to promote generalization," Journal of Speech and Hearing Disorders, vol. 53, no. 4, pp. 392–399, 1988.
4. D. Edgar and L. Rosa-Lugo, "The critical shortage of speech-language pathologists in the public-school setting: Features of the work environment that affect recruitment and retention," Language, Speech, and Hearing Services in Schools, vol. 38, no. 1, pp. 31–46, 2007.
5. Boyce, S. (2015). The Articulatory Phonetics of /r/ for Residual Speech Errors. *Seminars in Speech and Language*, *36*(04), 257–270. https://doi.org/10.1055/s-0035-1562909
6. Twist, A., Baker, A., Mielke, J., and Archangeli, D. (2007). "Are 'covert' // allophones really indistinguishable?" University of Pennsylvania Working Papers in Linguistics 13(2), 207–216.
7. P. Delattre and D. C. Freeman, "A Dialect Study Of American R's By X-Ray Motion Picture," Linguistics, vol. 6, no. 44, 1968.
8. C. Espy-Wilson and S. Boyce, "Acoustic differences between 'bunched' and 'retroflex' variants of American English /r/," Journal of the Acoustic Society of America, vol. 95, p. 2823, 1994.
9. X. Zhou et al., "A magnetic resonance imaging-based articulatory and acoustic study of 'retroflex' and 'bunched' American English /r/," JASA, vol. 123, no. 6, p. 16, 2008.
10. Westbury, J. R., Hashi, M., and Lindstrom, M. J. (1998). "Differences among speakers in lingual articulation for American English //," Speech Commun. 26, 203–226.
11. Tiede, M. K., Boyce, S. E., Holland, C. K., & Choe, K. A. (2004). A new taxonomy of American English /r/ using MRI and ultrasound. *The Journal of the Acoustical Society of America*, *115*(5), 2633–2634. https://doi.org/10.1121/1.4784878
12. H. King and E. Ferragne, "Loose lips and tongue tips: The central role of the /r/−typical labial gesture in Anglo-English," Journal of Phonetics, vol. 80, Article 100978, 2020.
13. P. Boersma and D. Weenink, "Praat: Doing phonetics by computer (Version 6.0.50)," 2019. [Online].
14. M. K. Tiede, "GetContours: Tongue contour fitting software (Version 3.3)," 2020.
15. C. Laporte, "SLURP. Version: 0.9," GitHub, 2018. [Online].
16. M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi," in Interspeech 2017, 2017, pp. 498–502.
17. N. R. Benway, J. L. Preston, E. Hitchcock, Y. Rose, A. Salekin, W. Liang, and T. McAllister, "Reproducible speech research with the artificial intelligence–ready PERCEPT corpora," Journal of Speech Language and Hearing Research, vol. 15, pp. 1–24, 2023.
18. H. Valentine, J. MacAuslan, M. Grigos, and M. Speights, "My Vowels Matter: Formant Automation Tools for Diverse Child Speech," in Proc. INTERSPEECH 2023, 2023, pp. 674-675.

This research was supported by NIH NIDCD F31DC022514 (PI: Eads), F31DC018197 (PI: Kabakoff), R01DC013668 (PI: Whalen), and R01DC017476 (PI: McAllister). Additional support provided by the American Speech-Language-Hearing Foundation, the Acoustical Society of America, and the Council of Academic Programs in Communication Sciences and Disorders.