Comparing tongue movement vs. shape representations from ultrasound imaging of /ɑɹ/ articulatory strategies

Sarah R. Li,[1] Sarah Dugan, PhD CCC-SLP,[2,3] Sarah M. Schwab, PhD,[4] Colin T. Annand, PhD,[4] Jack Masterson,[1] Reneé Seward, MGD,[5] Suzanne Boyce, PhD CCC-SLP,[2] Michael A. Riley, PhD,[3] & T. Douglas Mast, PhD[1]

[1]Biomedical Engineering; [2]Communication Sciences and Disorders; [3]Rehabilitation, Exercise, and Nutrition Sciences; [4]Psychology; [5]Design | University of Cincinnati

University of Cincinnati

Introduction

- **Redundancy** in motor control systems allows for multiple underlying solutions that cause relatively invariant functional outcomes.
  - One such outcome is the transition between articulated phonemes in speech sound production [1, 2].
- **Ultrasound imaging** provides real-time visual biofeedback of the tongue, which may be used to refine movements towards a specific solution (strategy used to produce a desired speech sound).
- Tongue **shapes** vs. **movements** in analyzing articulatory strategies:
  - Many studies of speech have focused on static positioning of articulators (i.e., tongue shapes) as endpoints of these transitions.
  - However, /ɹ/ tongue shapes vary due to adjacent vowels [3, 4].
  - As endpoints, tongue shapes may be representative of transitions; conversely, transitions may differ as another form of redundancy.

Methods

Tongue part displacement trajectories

- Automatic tongue movement measurement via TonguePART [4, 5]
- Quantifies time-dependent vocal tract constrictions due to the root, dorsum, and blade (positive displacement = narrowing vocal tract).

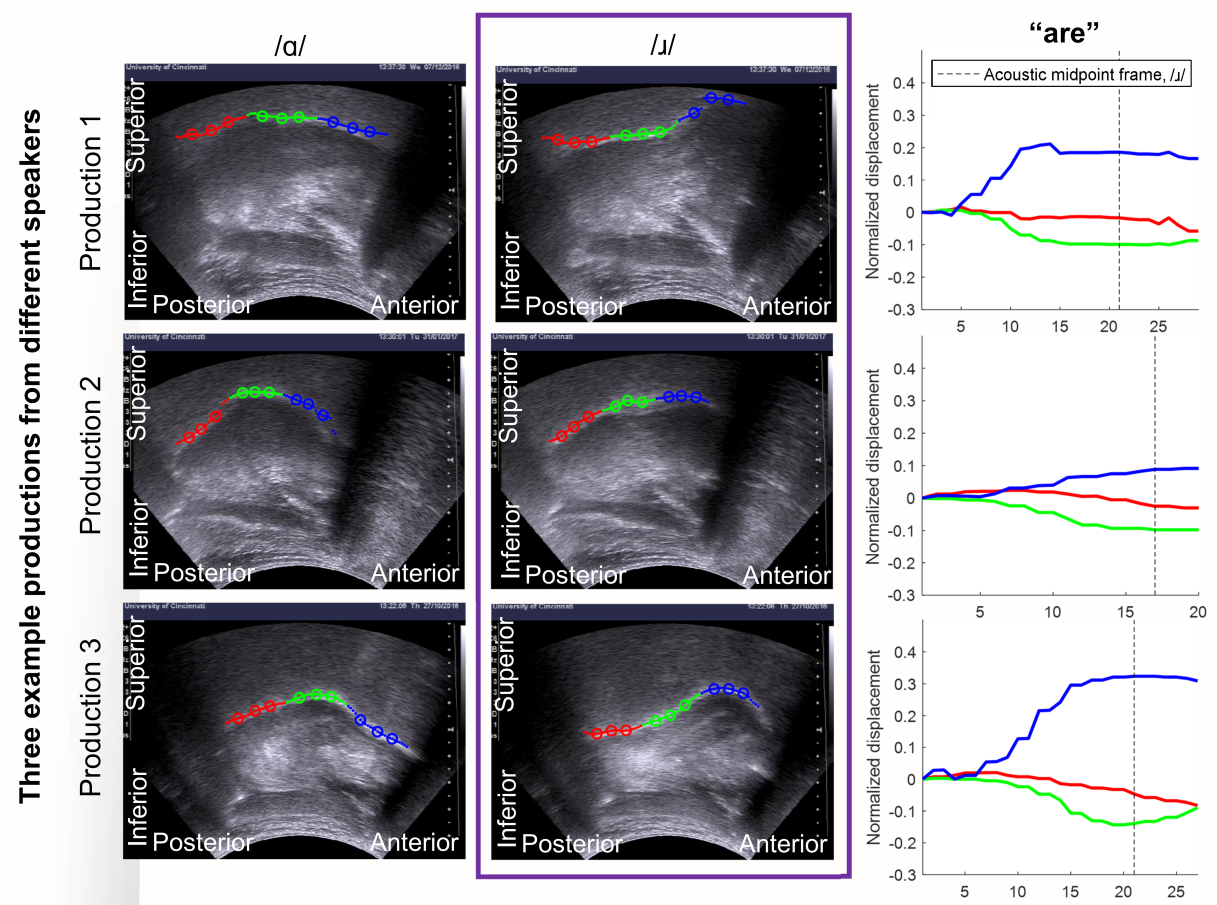


Figure 1: Tracked ultrasound images at the acoustic midpoints of /ɑ/ (left column) and /ɹ/ (middle column) from three /ɑɹ/ ("are") productions by different speakers, with measured displacement trajectories (right column).

Data representations

- Mid-sagittal ultrasound image sequences of /ɑɹ/ were recorded from 27 typical adult speakers of American English (399 productions total).
- Measures of tongue shape (A) and movement (B, C, D) considered, with (C) and (D) as lower-dimensional representations of (B):

(A) Normalized vertical and horizontal coordinates of the tongue contour at the midpoint-/ɹ/ frame (as seen before normalization in Fig. 1, middle column).
(B) Tongue part displacement trajectories (as seen in Fig. 1, right column).
(C) Scores from principal component analysis (PCA) of tongue part displacement trajectories. PCA is a linear reduction of dimensions via projection in directions of greatest variance. Three PCs were used for analyses (two PCs are used for 2D visualization in Figs. 6, 7, and 8).
(D) Displacements of root, dorsum, and blade between midpoint-/ɑ/ and midpoint-/ɹ/ frames (e.g., at dashed vertical lines in Fig. 1, right column).

Methods (cont'd)

Analyses of representations

- To compare articulatory information provided:
  1. Euclidean distances were calculated between all possible pairs of productions.
  2. R^2 values from cross-correlation of distances quantified similarity.
     - Comparable to assessing the fit of distances in multi-dimensional scaling [6].
- To compare identification of /ɹ/ articulation strategies:
  - k-means clustering (k=2, from gap criterion) found clusters that minimized distances between data points and centers of their containing clusters.
  - Jaccard indices show cluster agreement (from 0 to 1: no similarity to exact similarity).

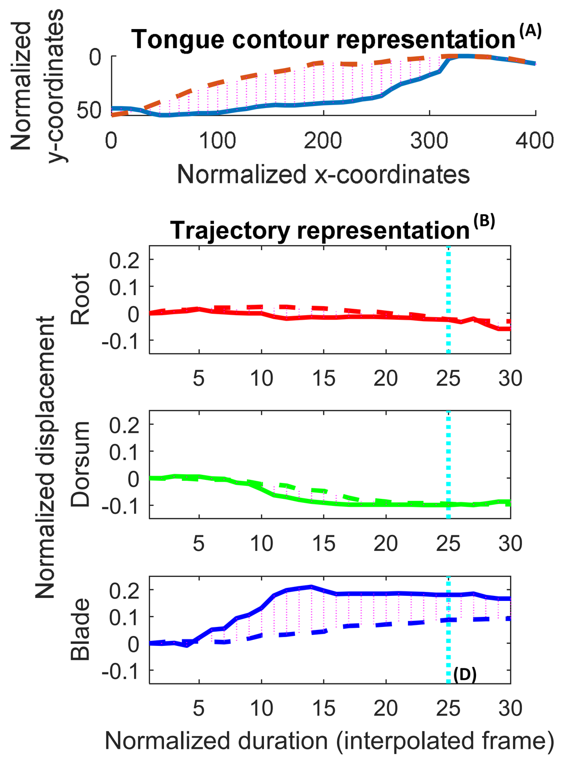
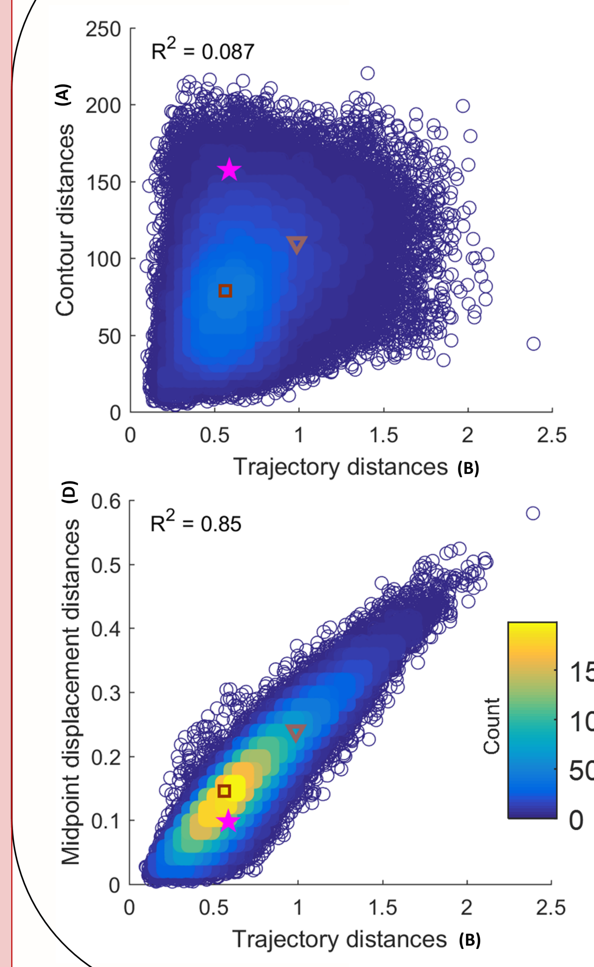


Figure 2: *To compare articulatory information provided:* Contour (top) and displacement trajectory (bottom) representations for Productions 1 and 2 from Fig. 1, with magenta dotted lines illustrating differences used for the Euclidean distance calculation. The relative-displacement representation is shown by vertical cyan dashed lines at the /ɹ/-midpoint frame. The three distances calculated for this pair in the contour (A), trajectory (B), and displacement at the /ɹ/-midpoint frame (D) representations are shown with magenta stars in Fig. 4 (A and D on the y-axes, B on the x-axes).

Results

Comparison of articulatory information



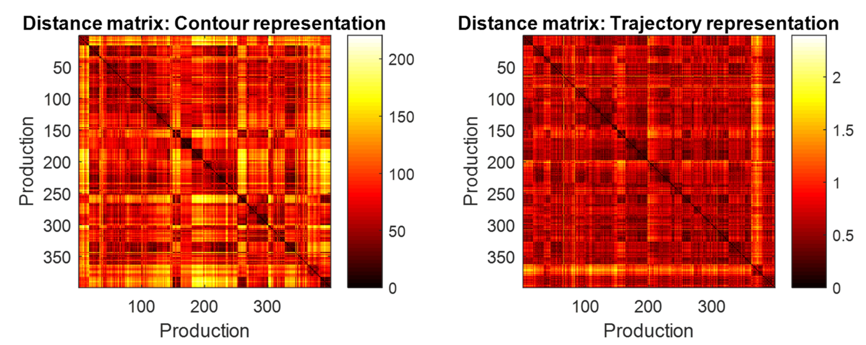


Figure 3: *(Above)* Euclidean distance matrices. Each matrix element shows Euclidean distances between two productions in the dataset for the representation used. Distances from these matrices are plotted vs. each other in Fig. 4 (top).

- See Table 1 for R^2 values (Fig. 4).

Figure 4: *(Left)* Euclidean distances between all recorded productions, comparing contour vs. trajectory (top) and relative-displacement at midpoint-/ɹ/ vs. trajectory (bottom) representations. A heatmap indicates the count of pairs overlapping in these scatterplots. Markers indicate distances for the example productions from Fig. 1.

Tables

Table 1: For the trajectory representation compared to other representations: R^2 values from cross-correlation of distances; Jaccard indices of cluster similarity for both clusters.

Data representation	R^2	Jaccard indices
Contours	0.087	0.40, 0.34
Scores (3 PCs, 81.7% variance explained)	0.94	0.99, 0.98
Displacements at /ɹ/	0.85	0.85, 0.84

Table 2: Cluster membership for the example productions shown in Fig. 1.

Example production	Cluster	
	Contour	Trajectory
Production 1	1	1
Production 2	2	1
Production 3	1	2

Comparison of clusters

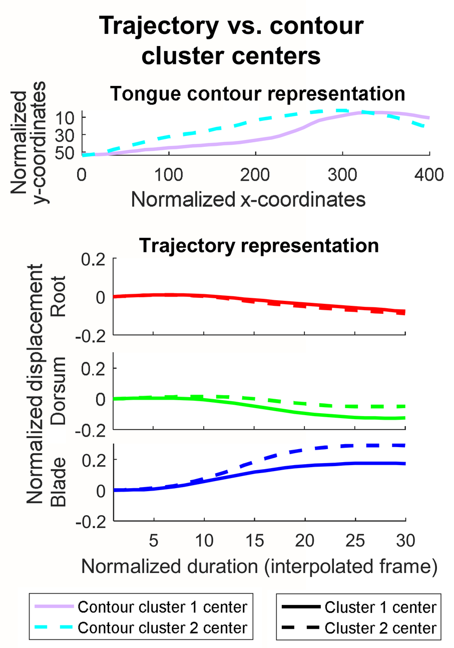


Figure 5: Contours (top) and trajectories (bottom) representing cluster centers for contour and trajectory representations (Fig. 6).

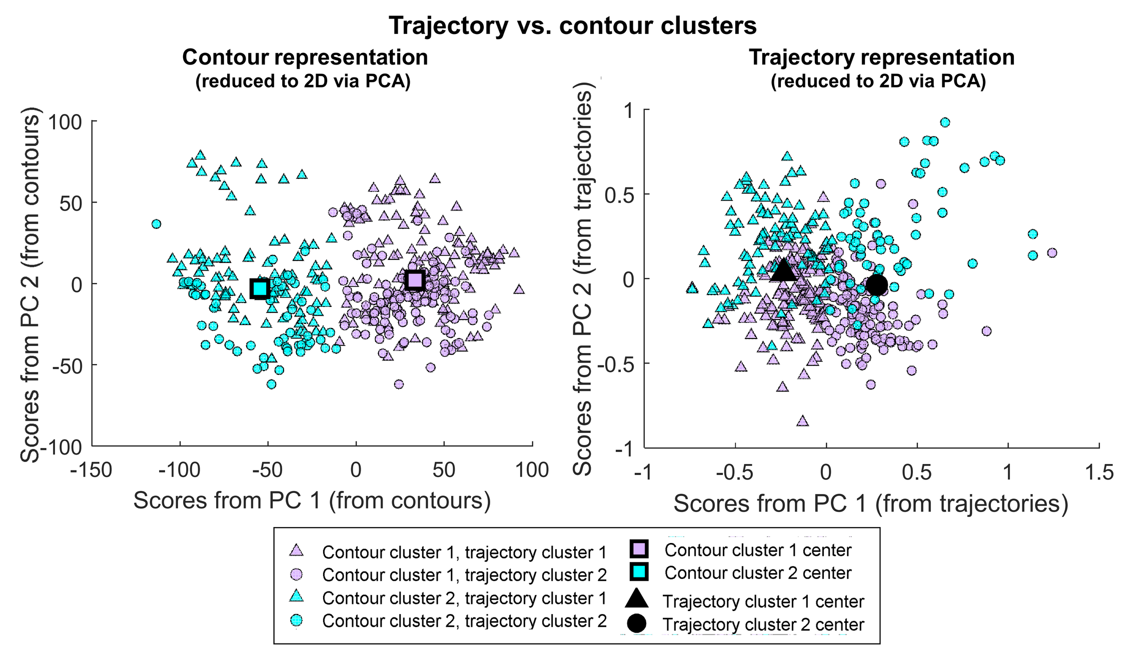


Figure 6: Cluster memberships (marker color indicating contour cluster membership; shape indicating trajectory cluster membership) plotted in 2D representations, using the first two principal components (PCs) from PCA on contour (left, 70.4% variance explained) and trajectory (right, 65.4% variance explained) representations. Markers indicate cluster centers (shown in Fig. 5 without dimension reduction).

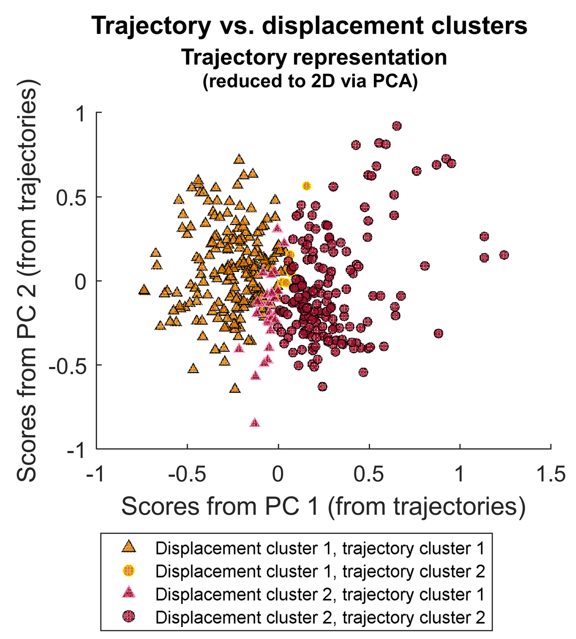


Figure 7: Cluster membership shown by the indicated markers for relative-displacement at /ɹ/ and trajectory representations.

- See Table 1 for Jaccard indices indicating membership similarity of clusters illustrated in Figs. 6 and 7.
- See Table 2 for cluster membership of example productions in Fig. 1 (compare to centers in Fig. 5).

Further use

- Separate dataset of three children (two with residual speech sound disorder, RSSD; one with typical speech) using a gamified ultrasound biofeedback display.
- RSSD speakers differed relative to typical speaker: one closer with contour, the other with trajectory representations (Fig. 8).
- May indicate different speech difficulties (movement direction vs. timing).

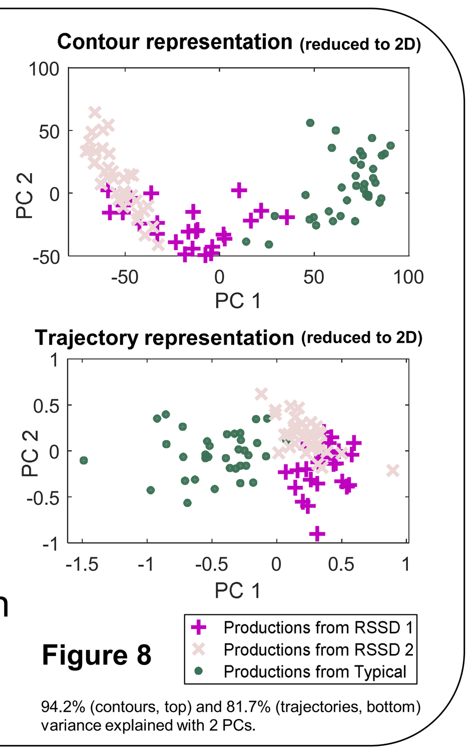


Figure 8

94.2% (contours, top) and 81.7% (trajectories, bottom) variance explained with 2 PCs.

Discussion and conclusion

- **Different articulatory information is present in contour and trajectory representations:**
  - Differences between distance matrices (Fig. 3).
  - Large spread in top panel of Fig. 4 (R^2 = 0.087).
  - Identified clusters differ greatly (Fig. 6; Jaccard indices ≤0.40 in Table 1).
- **Lower-dimensional representations capture most of the variance in trajectories:**
  - Narrow spread in the bottom panel of Fig. 4.
  - High R^2 values (0.85 and 0.94).
  - Identified clusters differ minimally (Fig. 7; Jaccard indices ≥0.84 in Table 1).
- **Identified clusters are separated by direction of concavity (contour) or blade displacement (trajectory)**, shown in Figs. 5 and 6.
- **Both contours and trajectories suggest a continuum of articulatory strategies:**
  - Spread in Fig. 6 (lack of separated groups).
- **These data representations can be used to compare tongue movement articulation in different productions.**
  - Example comparison of articulatory strategies: Productions 1 vs. 3 are closer (smaller distances) than Productions 2 vs. 3. Productions 1 vs. 2 are close by trajectories but not by contours. (Fig. 4; cluster memberships in Table 2.)
  - Typical vs. RSSD speech (Fig. 8).

Acknowledgements

- NIH/NIDCD grants R01 DC017301 and R01 DC013668.
- University of Cincinnati Creating our Third Century Funding Support.
- Siemens Medical Solutions for lending the Acuson X300 ultrasound scanner.

References

[1] Tiede, M. K., Boyce, S. E., Holland, C. K., & Choe, K. A. (2004). A new taxonomy of American English /r/ using MRI and ultrasound. *J. Acoust. Soc. Am.*, 115(5), 2633–2634.
[2] Westbury, J. R., Hashi, M., & Lindstrom, M. J. (1998). Differences among speakers in lingual articulation for American English /ɹ/. *Speech Commun., 26*(3), 203–226. https://doi.org/10.1016/S0167-6393(98)00058-2
[3] Mielke, J., Baker, A., & Archangeli, D. (2010). Variability and homogeneity in American English /r/ allophony and /s/ retraction. In *Variation, Detail, and Representation*. (pp. 699–719). Mouton de Gruyter.
[4] 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. *Perspect. ASHA Spec. Interest Groups, 4*(6), 1644–1652. https://doi.org/10.1044/2019_PERS-19-00064
[5] Li, S. R., Dugan, S., Masterson, J., Hudepohl, H., Annand, C., Spencer, C., Seward, R., Riley, M. A., Boyce, S., & Mast, T. D. (2022). Classification of accurate and misarticulated /ɑr/ for ultrasound biofeedback using tongue part displacement trajectories. *Clin. Linguist. Phon.*, 1–27. https://doi.org/10.1080/02699206.2022.2039777
[6] Wang, J., Green, J. R., Samal, A., & Yunusova, Y. (2013). Articulatory distinctiveness of vowels and consonants: A data-driven approach. *J. Speech Lang. Hear. Res., 56*(5), 1539–1551. https://doi.org/10.1044/1092-4388(2013/12-0030)