Discrete constriction locations describe a comprehensive range of simulated vocal tract shapes

JL Gaines[1], KS Kim[2], B Parrell[3], V Ramanarayanan[2,4], SS Nagarajan[2], JF Houde[2]

[1]UC Berkeley-UCSF Graduate Program in Bioengineering, [2]Department of Otolaryngology, University of California, San Francisco, [3] Department of Communication Sciences and Disorders, University of Wisconsin–Madison, [4]Educational Testing Service R&D, San Francisco

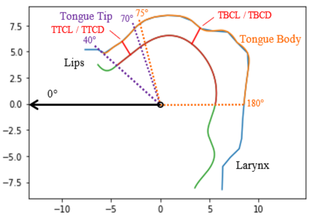
Introduction

At 10-15 phonetic segments per second, speech is one of the fastest-paced motor tasks that humans perform. Characterizing this task is relevant in understanding a variety of speech and motor control disorders. Here we sought to explore patterns across all vocal tract constrictions used to produce vowel sounds.

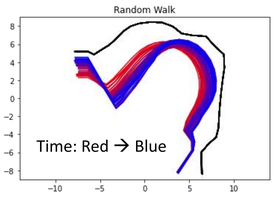
The Maeda model of the vocal tract [1] was used to synthesize the vowel sounds resulting from a comprehensive range of vocal tract shapes. An exploration of the dataset revealed that constriction in the tongue body region was limited to three discrete regions, despite the dataset covering a complete range of formant values.

Methods

- The Maeda model [1] is based on a statistical analysis of vocal tract movements, resulting in seven principal components (PCs) that correspond strongly to articulator movements (e.g., lip opening, tongue position, tongue shape)
  - Each PC takes values from -3 to 3 standard deviations (sd) from the mean
- Data for this exploration was generated as a series of random walks through the Maeda model PC space
  - The PC corresponding to the larynx was held at 0
  - As a starting point, each of the six remaining PCs was assigned a random value between -3 and 3 sd
  - If the starting point was valid, a random step was applied to find the next point
  - A valid vocal tract shape was defined to have five formants and F1 between 250 and 900 Hz
  - For each dimension, a step size drawn from a uniform random distribution between -0.25 and 0.25 sd was applied to the current vocal tract shape until the shape was no longer valid or the random walk reached a maximum of 50 steps
  - Then a new starting point was randomly selected for the next walk
- Finally, the following parameters were calculated from each vocal tract shape:
  - The location and degree of the smallest constriction at the tongue body (75° to 180° from horizontal)
  - The location and degree of the smallest constriction at the tongue tip (40° to 70° from horizontal)

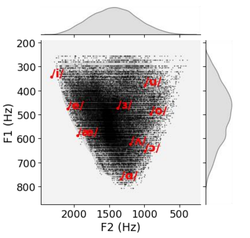


Example Random Walk



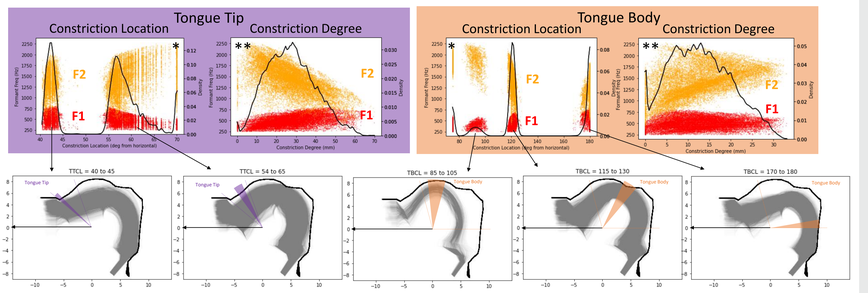
Formant Space

- The dataset covers the F1-F2 space, showing that a full range of vowel sounds can be produced with the vocal tract shapes included in the dataset
- For reference, average formant frequencies of different vowel sounds are plotted for male American English speakers [2].
- This result provides evidence that the dataset contains vocal tract shapes needed to produce a full range of speech sounds.



Constriction Location and Degree

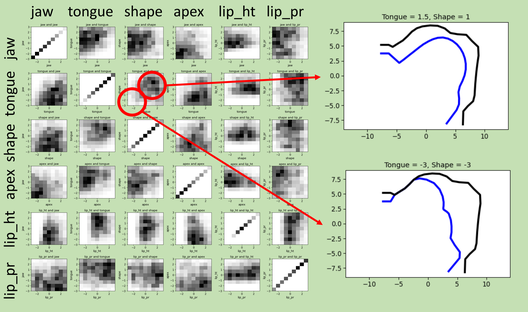
- Constrictions in the tongue body are restricted to three discrete locations
- Constrictions in the tongue tip are less restricted but show higher density in two regions
- F1 (red) and F2 (orange) frequencies are plotted for each vocal tract shape
- The subset of vocal tract shapes corresponding with each constriction location are plotted. Darker regions indicate greater density of observations.



* Indicates a boundary artifact. If there is no constriction in the region, the narrowest point will be at the boundary between regions.
** Vocal tract shapes that produced fewer than five formants were excluded in order to focus on vowel productions. This may explain the low density of observations with small constriction degrees.

Side note: Two-dimensional associations

- Since random walks were terminated when an invalid shape was reached, the data set is made up of only valid vocal tract shapes (those that produce five nonzero formants and F1 between 250 and 900 Hz)
- Two-dimensional histograms can show which PC values commonly co-occur in valid vocal tract shapes, and which combinations commonly lead to invalid shapes.
- For example, negative values of tongue dorsal position “tongue”, corresponding with positions closer to the front of the mouth, are often invalid when combined with negative values of tongue shape “shape”, which correspond with a flatter tongue. If the tongue is flat, it will collide with the alveolar ridge at more frontal positions, leading to an invalid vocal tract shape for vowel production.



Discussion

- Coverage of the two-dimensional F1-F2 space was demonstrated. Coverage of higher dimensional formant space is more difficult to demonstrate due to the curse of dimensionality; however, coverage of the F1-F2 space ensures that all vowel sounds are present in the data set.
- The Maeda model does not reflect the kinematics of a moving vocal tract, nor the physics of air pressure or velocity, so this exploration is limited to still-frame data
- The Maeda model was created from data collected from female French speakers, then transformed to reflect a typical male vocal tract. Thus the vocal tract shapes discussed in this exploration are not directly observed from a male speaker.
- Boë et al [3] found that each of 10 French vowels appear to have a characteristic set of discrete tongue constriction locations. This exploration suggests that a comprehensive range of vowel-producing vocal tract shapes can be reduced to a total of three discrete constriction locations in the tongue body.

References

[1] S.Maeda. In *Speech production and speech modeling,* W. Hardcastle & A. Marchal, Eds. The Netherlands: Kluwer Academic Publishers, 1990, pp. 131-149.

- Maeda model code was obtained from the GitHub of Satrajit Ghosh (https://github.com/satra/VocalTractModels)

[2] J.Hillenbrand, et al. *J Acoust Soc Am.* 1995, 97(5), 3099. doi:10.1121/1.411872.

[3] L.-J.Boe, et al. *J Phon.* vol. 20, no. 1, pp. 27-38, 1992. https://doi.org/10.1016/S0095-4470(19)30251-7.