Which cues do listeners use? Discovering networks of phonetic cues for speech sound categorization using a graph theoretic approach

Anne Marie Crinnion
acrinnion@college.harvard.edu
Harvard University

Beth Malmskog
beth.malmskog@villanova.edu
Villanova University

Joe Toscano
joseph.toscano@villanova.edu
Villanova University

INTRODUCTION

Problem:

- Human speech is highly variable, yet listeners have little difficulty recognizing specific phonemes [1]
- Cue-integration models have been proposed as a solution [2,3] but they require assumptions about what the relevant cues are (based on phonetic measures of hypothesized cues)
- Need methods to isolate cues that are used consistently across large groups of listeners

Proposed solution:

- Use methods from graph theory to understand listener classification
- Steiner Tree algorithm used to identify subgraphs that minimize edge weights while connecting relevant nodes [4]
- Acoustic measurements and fricative classification data used to evaluate model [3,5]

METHOD

Acoustic data:

- 23 acoustic cue measurements across eight fricatives (/f,v,θ,ð,s,z,ʃ,ʒ /)
- Created bins spanning range of acoustic values for each cue
- Code assigned to each bin, creating 230 possible cue-value combinations

Graph structure:

- Nodes: 20 listeners and 230 codes
- Edges: weighted by inverse likelihood of response probability for a specific fricative given the code

Sub-graph search procedure:

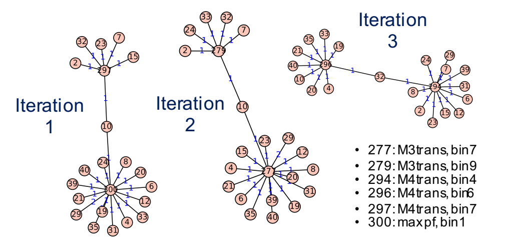
- Subgraphs calculated in SageMath
- After first solution found, identified cue-value nodes removed, and algorithm run again; repeated until no remaining cue-values form a graph connecting *all* listeners

RESULTS

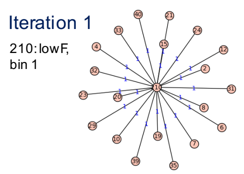
Representative Steiner Tree solutions

/f/ Steiner Trees

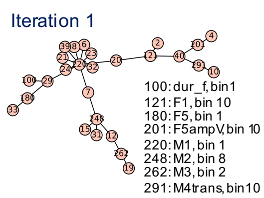
2-digit nodes: Listeners; 3-digit nodes: Cues



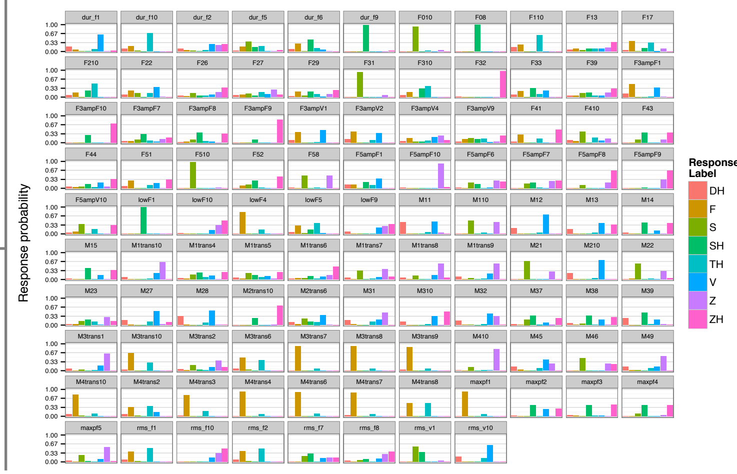
/ʃ/ Steiner Tree



/ð/ Steiner Tree



Phoneme identification probabilities for cues discovered by algorithm



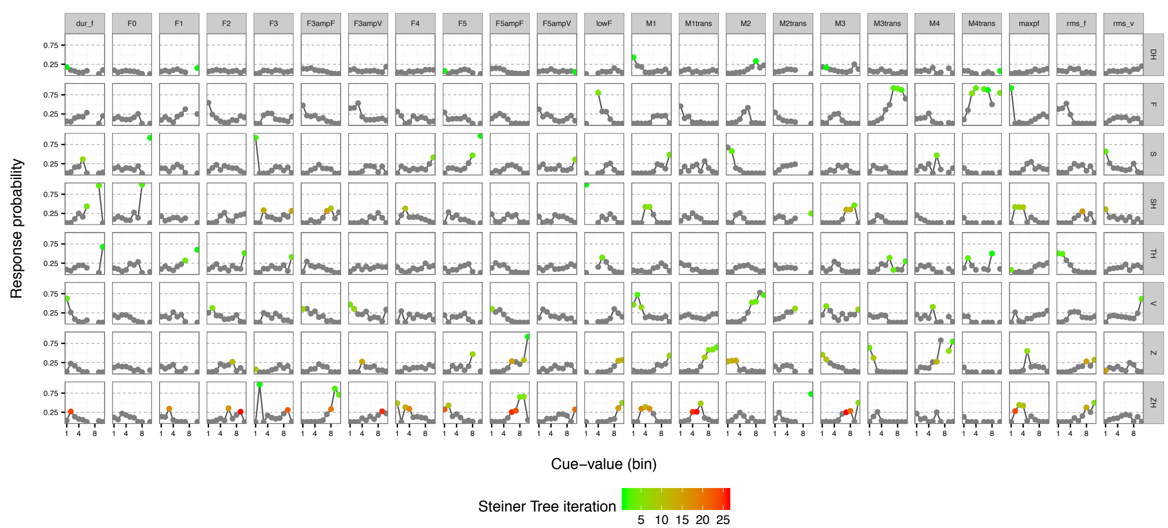
Steiner Tree structures:

- Some phonemes are highly robust, with multiple cues that connect all listeners on a single run (e.g., /ʃ/)
- Other phonemes have no single cue used by all listeners, resulting in complex graphs (e.g., /ð/)

Model accuracy:

- Many cues correspond to 1-2 unique phonemes, suggesting that they are robust cues
- Some cues provide information about specific phonological features, but not necessary unique phonemes (e.g., F5 amplitude, bin 9: voicing)

Phoneme identification probability by cue-value (all cues in dataset)



DISCUSSION

- 107 unique cues connected all listeners
- Many cues are relevant in sound identification—supports massive cue-integration as an model of human speech perception
- Next step: Develop a model that mimics uses these cues to classify new sounds; measure cues, group into bins, map onto points in 107-dimensional space, and compute distance to each phoneme based on Steiner Tree solutions

REFERENCES

[1] Toscano, J.C., & Allen, J.B. (2014). Across and within consonant errors for isolated syllables in noise. *J Speech Lang Hear Res, 57*, 2293-2307.

[2] Toscano, J.C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cog Sci, 34*, 434-464.

[3] McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psycol Rev, 118*, 219-246.

[4] Sadeghi, A., & Fröhlich, H. (2013). Steiner tree methods for optimal sub-network identification: an empirical study. *BMC Bioinformatics, 14*.

[5] Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *J Acoust Soc Am, 108*, 1252-1263.