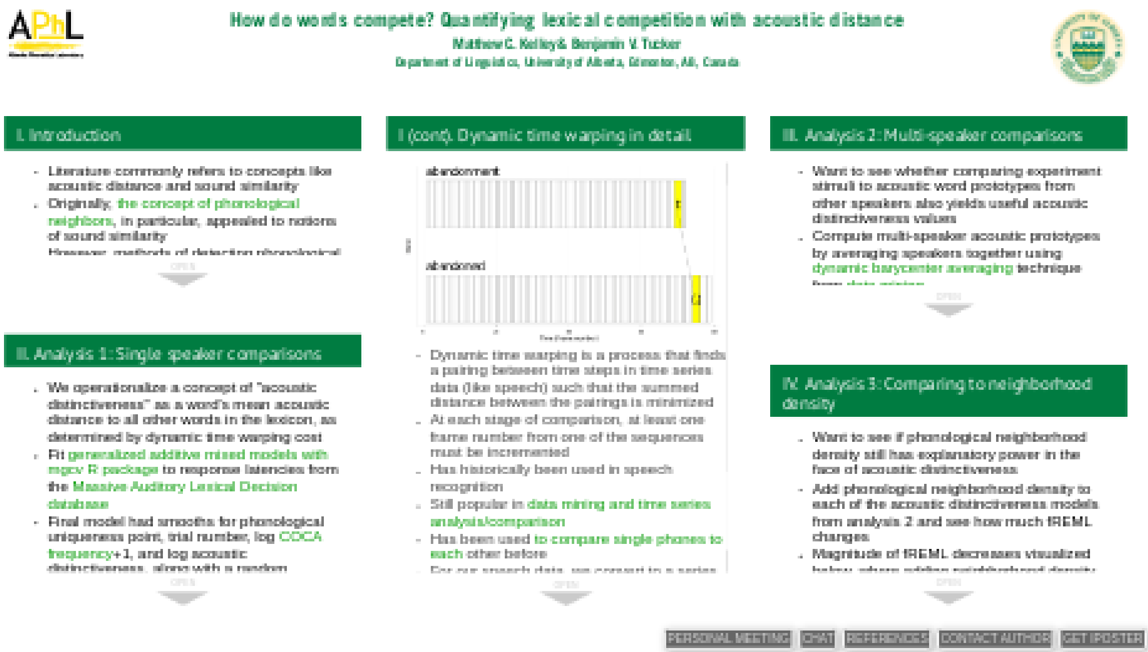
How do words compete? Quantifying lexical competition with acoustic distance



Matthew C. Kelley & Benjamin V. Tucker

Department of Linguistics, University of Alberta, Edmonton, AB, Canada

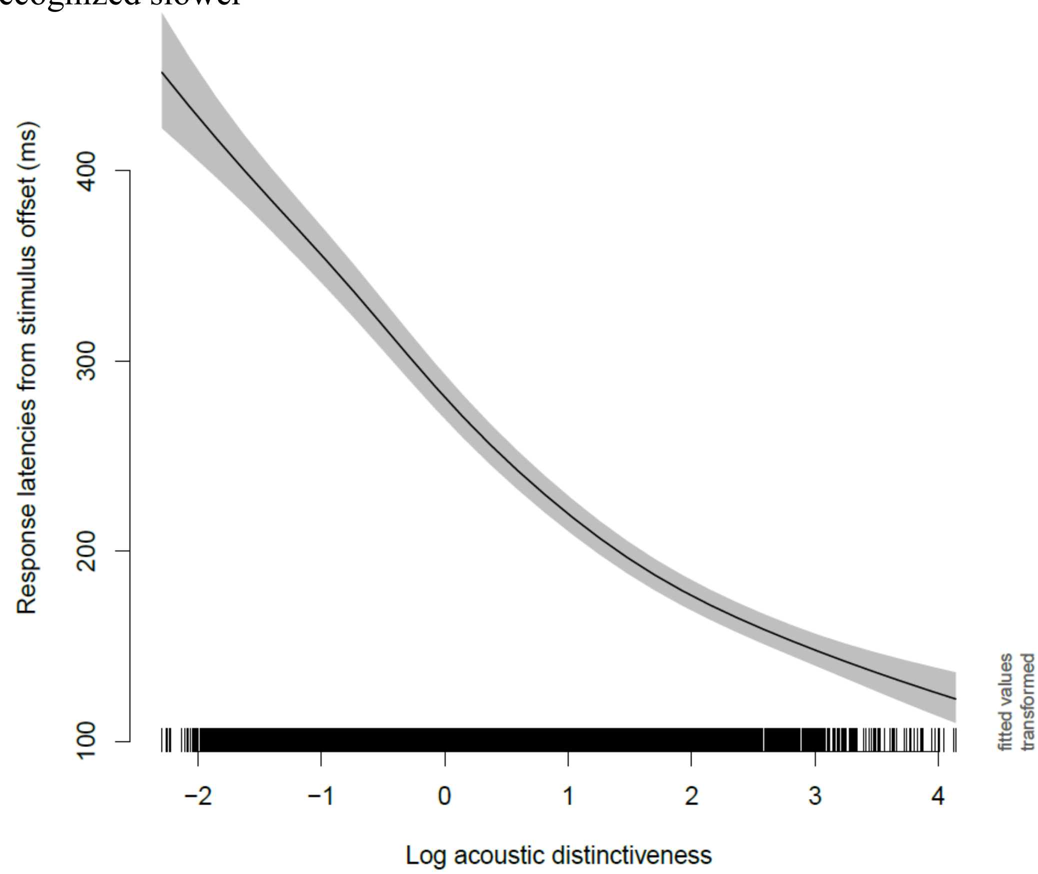
PRESENTED AT:

I. INTRODUCTION

- Literature commonly refers to concepts like acoustic distance and sound similarity
- Originally, the concept of phonological neighbors (https://eric.ed.gov/?id=ED353610), in particular, appealed to notions of sound similarity
- However, methods of detecting phonological neighbors have relied on edit distance for strings of symbols, rather than acoustic comparisons
- Acoustic distance between words has yet to be well defined
- Here, we propose using dynamic time warping cost from time-series analysis as a form of acoustic distance (see dynamic time warping panel for more details)
- **Hypothesis**: words that are more acoustically similar to other words will take longer to recognize, and words that are less acoustically similar to other words will take less time to recognize

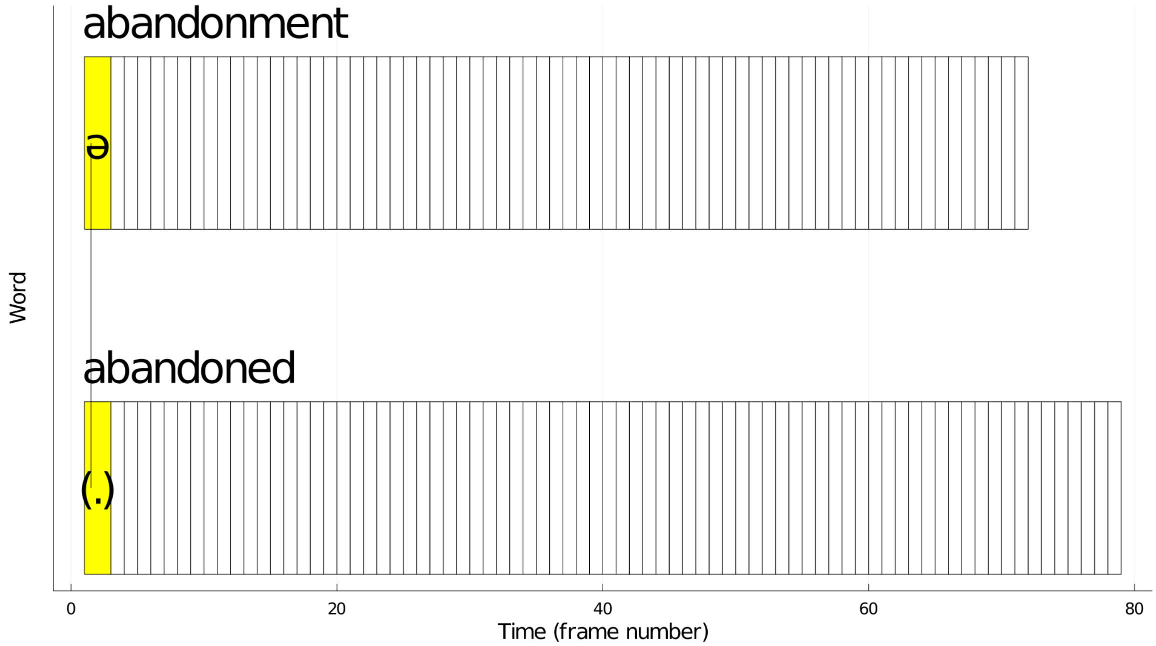
II. ANALYSIS 1: SINGLE SPEAKER COMPARISONS

- We operationalize a concept of "acoustic distinctiveness" as a word's mean acoustic distance to all other words in the lexicon, as determined by dynamic time warping cost
- Fit generalized additive mixed models with mgcv R package (https://doi.org/10.1111/j.1467-9868.2010.00749.) to response latencies from the Massive Auditory Lexical Decision database (https://doi.org/10.3758/s13428-018-1056-1)
- Final model had smooths for phonological uniqueness point, trial number, log COCA frequency (https://www.english-corpora.org/coca/)+1, and log acoustic distinctiveness, along with a random intercept for subject
- Smooth for log acoustic distinctiveness shown below. Trend is as expected, with words that are highly distinctive (have few competitors) are recognized faster, and words that are not very distinctive (have more competitors) are recognized slower



- Acoustic distinctiveness and acoustic distance seem to bear some relation to spoken word recognition overall and serve as evidence in favor of our hypothesis

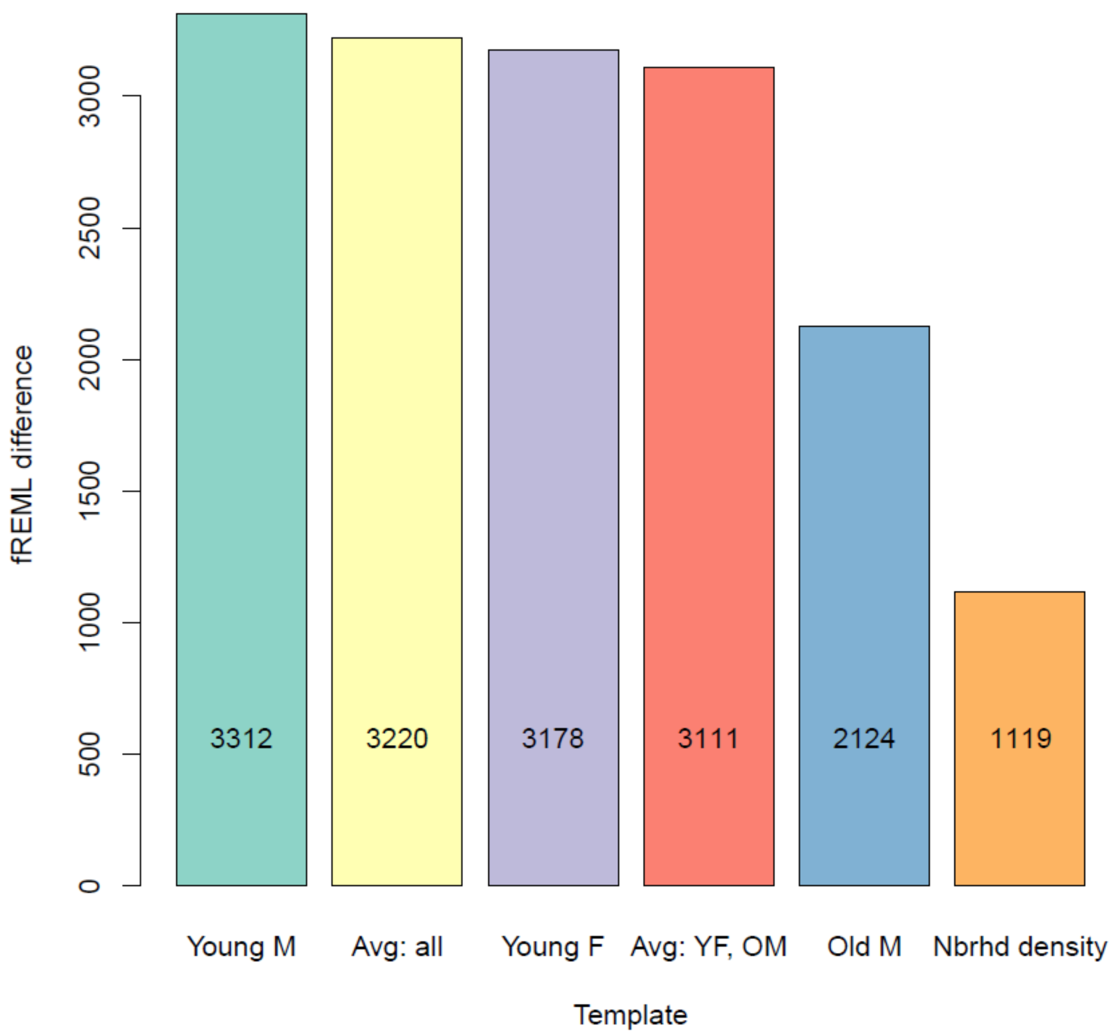
I (CONT). DYNAMIC TIME WARPING IN DETAIL



- Dynamic time warping is a process that finds a pairing between time steps in time series data (like speech) such that the summed distance between the pairings is minimized
- At each stage of comparison, at least one frame number from one of the sequences must be incremented
- Has historically been used in speech recognition
- Still popular in data mining and time series analysis/comparison (https://doi.org/10.1145/2339530.2339576)
- Has been used to compare single phones to each (https://doi.org/10.1016/j.lingua.2011.04.006) other before
- For our speech data, we convert to a series of Mel frequency cepstral coefficient vectors, calculated over 25 ms windows spaced at 10 ms intervals
- Opting not to use delta and delta-delta features (first- and second-order discrete derivatives) for now to simplify analysis

III. ANALYSIS 2: MULTI-SPEAKER COMPARISONS

- Want to see whether comparing experiment stimuli to acoustic word prototypes from other speakers also yields useful acoustic distinctiveness values
- Compute multi-speaker acoustic prototypes by averaging speakers together using dynamic barycenter averaging (https://doi.org/10.1016/j.patcog.2010.09.013) technique from data mining (https://doi.org/10.1109/ICDM.2014.27)
- The three speakers were the young male speaker from the Massive Auditory Lexical Decision database, in addition to a young female speaker and an older male speaker reading the same words
- Compare models with different acoustic distinctiveness calculations to baseline model using fREML from R itsadug package (https://cran.r-project.org/web/packages/itsadug/index.html), where greater reductions in fREML indicate better fit
- All models fit from the baseline model caused significant reductions in fREML, and the magnitudes of the reductions are visualized in the figure below

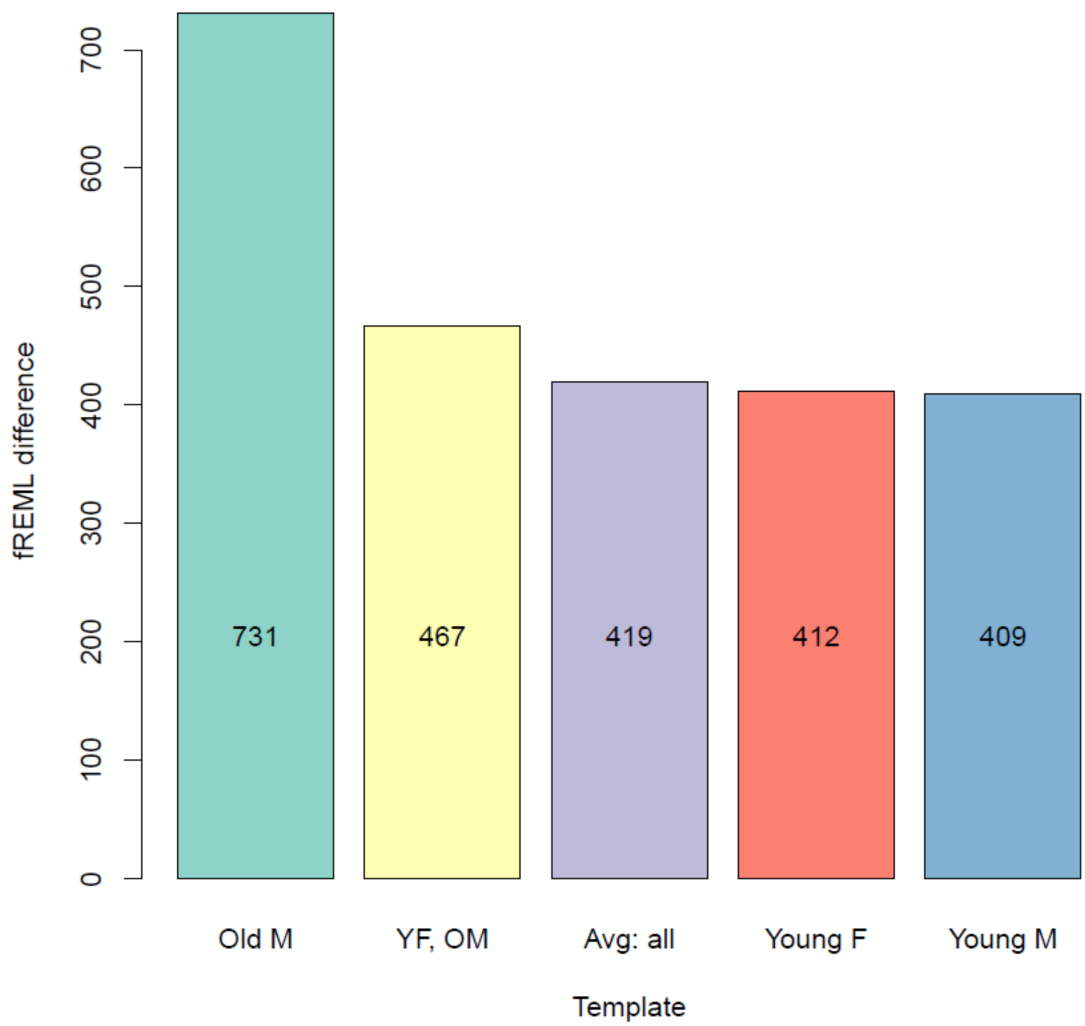


- Prototypes that included the young male speaker caused greatest decrease, likely due to containing acoustic information the participants heard (similar to evaluating a machine learning model on data it was trained on)

- All models using acoustic distinctiveness caused significantly greater decreases to fREML than neighborhood density did
- Results suggest that acoustic distinctiveness has a stronger relationship to the lexical decision data than does neighborhood density

IV. ANALYSIS 3: COMPARING TO NEIGHBORHOOD DENSITY

- Want to see if phonological neighborhood density still has explanatory power in the face of acoustic distinctiveness
- Add phonological neighborhood density to each of the acoustic distinctiveness models from analysis 2 and see how much fREML changes
- Magnitude of fREML decreases visualized below, where adding neighborhood density caused a decrease in for all models



- Overall, suggests that neighborhood density and acoustic distinctiveness are not measuring the exact same thing
- Question remains of how much of the remaining effect of neighborhood density has to do with lexical competition.
- For example, It is possible that a portion of neighborhood density's effect has to do with effects of orthography because phonological forms of English words are somewhat related to orthography
- Further experimentation is needed to determine whether neighborhood density is still useful in the face of a measure like acoustic distinctiveness

REFERENCES

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International journal of corpus linguistics, 14*(2), 159-190.

Luce, P. A. (1986). Neighborhoods of Words in the Mental Lexicon. Research on Speech Perception. Technical Report No. 6.

Mielke, J. (2012). A phonetically based metric of sound similarity. *Lingua, 122*(2), 145-163.

Petitjean, F., Ketterlin, A., & Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition, 44*(3), 678-693.

Petitjean, F., Forestier, G., Webb, G. I., Nicholson, A. E., Chen, Y., & Keogh, E. (2014, December). Dynamic time warping averaging of time series allows faster and more accurate classification. In *2014 IEEE international conference on data mining* (pp. 470-479). IEEE.

Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., ... & Keogh, E. (2012, August). Searching and mining trillions of time series subsequences under dynamic time warping. In *Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 262-270).

Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić, F., & Sims, M. (2019). The massive auditory lexical decision (MALD) database. *Behavior research methods, 51*(3), 1187-1204.

van Rij, J., Wieling, M., Baayen, R. H., and van Rijn, H. (2017). itsadug: Interpreting time series and autocorrelated data using GAMMs. R package version 2.3.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73*(1), 3–36.