Perception and timing of acoustic distance

Matthew C. Kelley[1] & Benjamin V. Tucker[2]
1. University of Washington
2. University of Alberta

1. Introduction

- Levenshtein distance commonly used to compare words, e.g., calculating neighborhood density [5]
- We proposed using dynamic time warping (DTW) to perform acoustic comparisons [2]
- DTW can be tweaked to better reflect cognition

Research questions

1. What distance function should be used in dynamic time warping? (Eq. 1)
2. How temporally elastic should the comparisons be? (Fig. 1)

2. Distance rating task

- DTW calculates distance on vectors like mel frequency cepstral coefficients (MFCCs)
- Monophthong vowels pairs synthesized with Praat's KlattGrid functionality using formants from [1]
- Participants rated how different a pair sounded on a scale of 1 to 7; ratings averaged across stimuli
- Acoustic distances (Eq. 1) calculated on MFCC representations of the vowels
- Correlation assessed between ratings and distances (Fig. 2)

$$d_p(x,y) = \left(\sum_i |\chi_i - \psi_i|^p\right)^{\frac{1}{p}}$$

Equation 1. The distance between two MFCC vectors x and y can be computed using the p-norm from linear algebra. Manhattan distance is calculated when $p = 1$, Euclidean distance when $p = 2$, and the maximum distance component when $p = \infty$.

- 2-norm performed well ($r = 0.878, p < 0.001$)
- Norms where $p \approx 4.5$ had highest correlation with distance ratings ($r = 0.883, p < 0.001$)

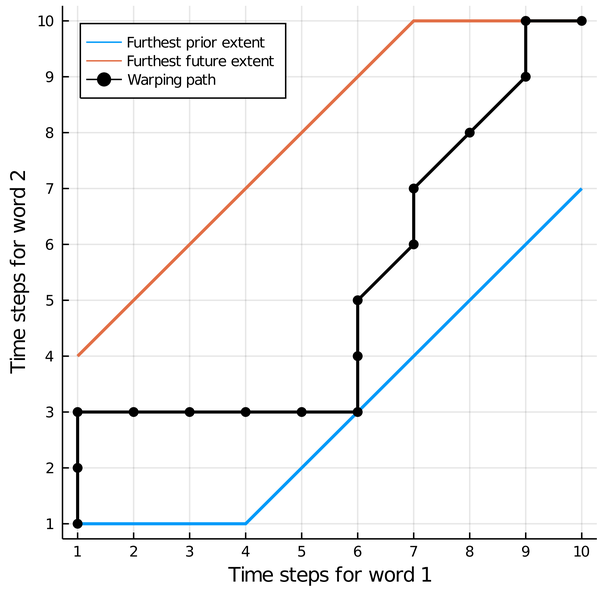

Figure 1. Tuning the elasticity of DTW constrains what time steps are allowed to be compared with each other.

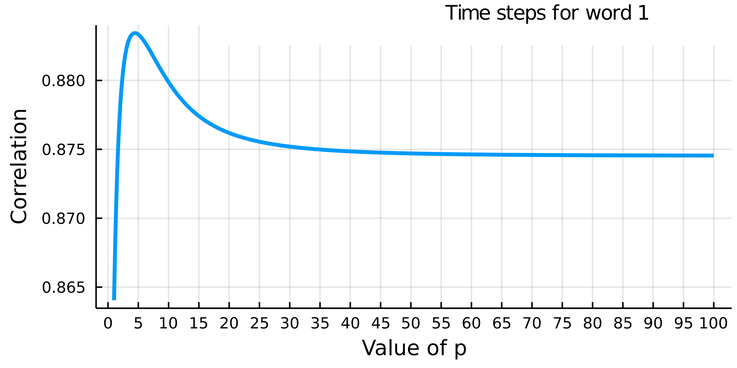

Figure 2. Correlations of distance functions with pooled participant distance ratings.

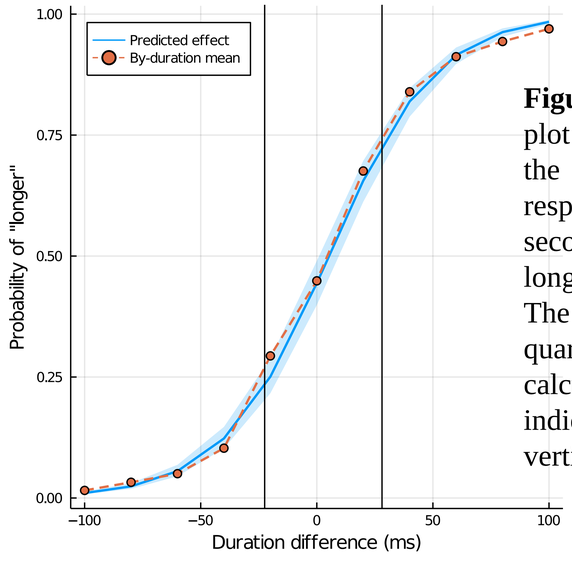

Figure 3. Effect plot of duration on the probability of responding that the second sound was longer than the first. The 1st and 3rd quartiles used in calculating JND are indicated with black vertical lines.

3. Duration discrimination task

- Can restrict how far in time a given time step can be stretched in DTW (Fig. 1)
- /i/, /ɑ/, /u/ vowels synthesized along a 200 ms continuum centered around average duration from [1]
- Reminder task: Center of duration continuum played, then the target stimulus
- Participant asked if second stimulus was longer or shorter than the first
- Logistic mixed-effects regression fitted to the data to estimate the just noticeable difference (JND) threshold for duration (Fig. 3)
- JND ≈ 25 ms, somewhat shorter than [3] and [4]

4. Behavioral reanalysis

- 4.5-norm and 25 ms JND were used in DTW to model response latencies in an auditory lexical decision task from [6]
- Repeated statistical analysis from [2]
- 2-norm and unconstrained elasticity induced better model fitness than experimental values
- Perception of acoustic distance can be reasonably modeled with many distance functions
- Temporal expectations in perception are tempered by acoustic distances

References:
[1] Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. The Journal of the Acoustical Society of America, 97(5), 3099–3111.
[2] Kelley, M. C., & Tucker, B. V. (2022). Using acoustic distance and acoustic absement to quantify lexical competition. *The Journal of the Acoustical Society of America*, *151*(2), 1367–1379.
[3] Kewley-Port, D., Watson, C. S., & Foyle, D. C. (1988). Auditory temporal acuity in relation to category boundaries; speech and nonspeech stimuli. The Journal of the Acoustical Society of America, 83(3), 1133–1145.
[4] Lapid, E., Ulrich, R., & Rammsayer, T. (2008). On estimating the difference limen in duration discrimination tasks: A comparison of the 2AFC and the reminder task. Perception & Psychophysics, 70(2), 291–305. https://doi.org/10.3758/PP.70.2.291
[5] Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spoken Words: The Neighborhood Activation Model. *Ear and Hearing*, *19*(1), 1–36.
[6] 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.

Contact: mattck@uw.edu

4pSC10, 182nd meeting of the Acoustical Society of America (Denver, CO) May 2022