# Comparing measures of word confusability and their effect on speech production

Esteban Buz & T. Florian Jaeger
Department of Brain and Cognitive Sciences University of Rochester

## 1. Introduction

- In the debate between **Ease of Production** and **Audience Design** accounts of word production an underlying issue is how to measure *confusability*
- Two widely discussed measures are **phonological neighborhood density (NHD), log frequency weighted neighborhood density (fNHD)**
- An underlying assumption of both is that *all phonemes are equally confusable* despite evidence that phonemes *differ in confusability* [1]
- Database (lexicon) choice is known to cause changes in effects [2]
- Effect of 'confusability' is known to differ from study to study with differing methods and measures [3, 4]

### Our Questions

1. How do NHD and fNHD compare to an a priori well defined measure of confusability (fCON) and how do each predict word production data?
2. Is the choice of lexical database important in creating these measures?
3. Why do studies differ in direction of effect for confusability in spoken production?

## 2. Calculating confusability

### Defining three measures of confusability

**NHD**

- For word *w*, NHD is the log count of all phonological neighbors
  - A phonological neighbor is one phoneme substitution away from the word in question (e.g. for bat: pat, bit, back)

**fNHD**

- For word *w*, fNHD is the sum of the log frequency of all phonological neighbors (with plus 1 smoothing)
  - Neighbor defined as above
  - $fNHD(w) = \sum_j \log(freq(n_j)+1)$, $n_j$ is *w*'s *j*-th neighbor

**fCON**

- For word *w*, fCON is the log probability of confusing that word with any phonological neighbor given phoneme to phoneme confusability and neighbor frequency (with plus 1 smoothing)
  - Neighbor defined as above
  - Adapted from the Luce & Pisoni frequency-weighted neighborhood probability rule [5, eq. 6]
  - Phoneme confusability is from the confusability matrices [1]

Equation 1. fCON detail. $n_j$ is the *j*-th neighbor of *w*, $PW_i$ is the *i*-th phoneme of *w*, $P(PW_i \mid PW_i)$ is the probability of identifying $PW_i$ when presented with $PW_i$

$$fCON(w) = 1 - \frac{freq(w) * \prod_i P(PW_i \mid PW_i)}{freq(w) * \prod_i P(PW_i \mid PW_i) + \sum_j [freq(n_j) * \prod_i P(PN_{ji} \mid PW_i)]}$$

### Lexical database choices

**CELEX2**

- 160,595 lexical entries with frequencies based on the COBUILD written English corpus [6]

**MRC2**

- 150,837 lexical entries with frequencies based on the Brown corpus (Kucera-Francis, [7])

**IPhOD2**

- 54,030 lexical entries with frequencies based on the SUBLEXus English movie subtitle corpus [8]

### Spoken word production data

- Log word durations taken from a picture naming experiment [9]
  - 40 targets; 36 participants
    - Excluded disfluent, mislabeled, outlier durations and targets with phonemes not found in the Woods et al (2010) data (/w/)
  - Targets are pairs of phonological neighbors (e.g. 'jar' & 'car)
    - Speakers uttered each target once, order of which target in a pair was spoken first was balanced across speakers

## 3. Measure comparisons

### Comparisons between and across databases

- Across databases, correlations on the above spoken word targets was high ($0.8 < r < 0.9$)

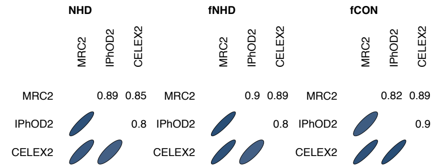



Figure 1. Across database comparisons

- Within database, correlations for NHD and fNHD were high ($0.86 < r < 0.92$) but neither was highly correlated with fCON ($0.31 < r < 0.47$)

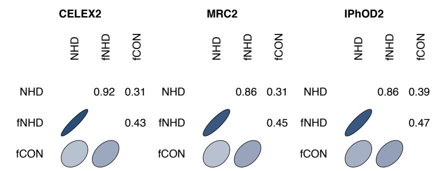



Figure 2. Within database comparisons

### Comparisons on spoken word data

- We use a LMM predicting spoken log word duration for each confusability measure
  - random intercepts by speaker and item
  - main effect of a given confusability measure
- All effects were negative, with and without controlling for log freq.
  - Significant effects for NHD in all databases, fNHD for CELEX2 & MRC2, fCON for IPhOD2
- Figure 3 shows example effects for the IPhOD2 database (log frequency and by speaker and item effects residualized out)

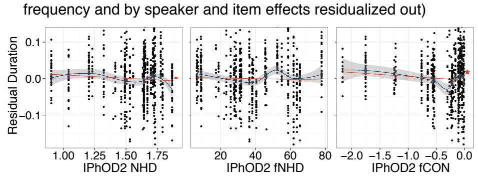



Figure 3. Confusability effect on residual log word duration, blue line is fit with LOESS smoother, red is fit with linear model (* = p < .05, . = p < .1)

## 4. Comparison with previous work

### Conflicting prior results and methods

- Many experimental studies bin items by confusability (high vs low)
  - These studies find positive correlations between word confusability and spoken duration [10, 3]
- Corpus studies typically use continuous measures of confusability
  - Results are mixed (negative effect in [4]; mixed effect in [9], Study 1)

### Within study comparison of binned vs. un-binned

- Each target was labeled as the high or low confusability item in its pair (e.g. for neighbor count in IPhOD2, 'jar' is low and 'car is high)
  - This followed the original design of [9]
  - Binning in this way holds constant much of the phonology across pairs

## 4. cont'd

### Results

- Binned confusability was only significant for IPhOD2 measures (this database was originally used to design the stimuli)
  - For all measures, the high bin *had longer durations*
  - For all measures, the effect direction for higher continuous confusability was negative
  - Frequency did not differ for NHD or fNHD bins ($t(14) < .65$) but did for fCON bins ($t(14) = -3.57$)
  - Figure 4 presents how measures relate in our data using IPhOD2

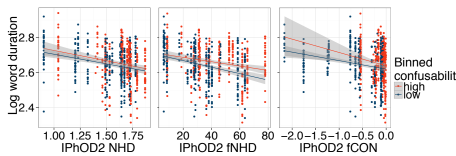



Figure 4. Effects of continuous confusability within each bin with best fit linear regression lines

## 5. Discussion

### Interpretation

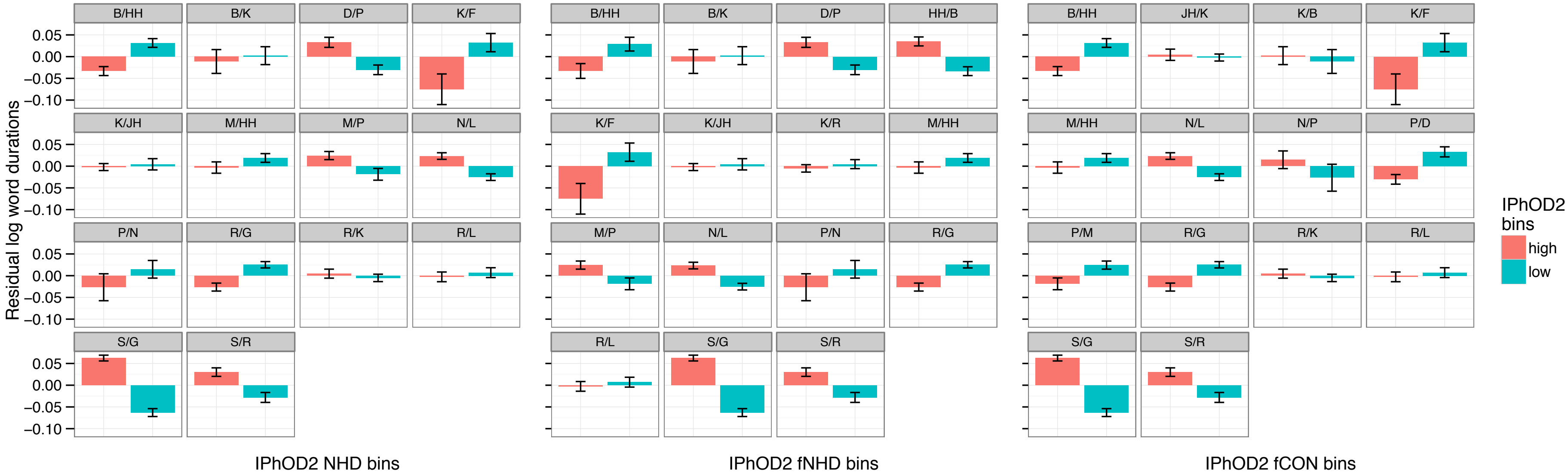
- Lexical database choice and measure can impact results
- fCON failed to produce much of an effect, however, some theoretical claims suggest that *any* effects of fCON are not predicted using this type of data and analysis [cf. 11]
  - phoneme-to-phoneme comparisons predict changes in phoneme articulation not necessarily spoken word duration
- The binned versus continuous effect difference suggests that usage of continuous confusability measures capture more than inter-word confusability
  - **Higher confusability predicts longer durations or no effect when phonology is held constant --> communicative/audience design accounts**
    - Our binned analysis compared how changes in confusability as the result of *a single phone difference* between words affects spoken duration
  - **Higher confusability predicts shorter durations or no effect when phonology is not held constant --> effects of "production ease" (cf. [4])**
    - Continuous measures of confusability such as NHD, fNHD and fCON may capture other properties of speech production

### Future questions

1. Do phoneme-specific measures of confusability predict phoneme-specific changes in spoken production?
2. Do these measures capture *intelligibility* in speech in addition to changes in speech rate?

**References**

1. Woods, D.L., et al., Consonant identification in consonant-vowel-consonant syllables in speech-spectrum noise. The Journal of the Acoustical Society of America, 2010. 127: p. 1609-23.
2. Yao, Y., The effects of phonological neighborhoods on pronunciation variation in conversational speech, in Department of Linguistics, 2011, University of California, Berkeley.
3. Scarborough, R., Lexical and contextual predictability: Confluent effects on the production of vowels, in Laboratory Phonology 10, C. Fougeron, et al., Editors. 2010, De Gruyter Mouton: Berlin; New York. p. 557-586.
4. Gahl, S., Y. Yao, and K. Johnson, Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. Journal of Memory and Language, 2012. 66(4): p. 789-806.
5. Luce, P.A. and D.B. Pisoni, Recognizing spoken words: the neighborhood activation model. Ear and hearing, 1998. 19: p. 1-36.
6. Baayen, R.H., R. Piepenbrock, and L. Gulikers, CELEX2, 1996, Linguistic Data Consortium: Philadelphia.
7. Wilson, M., MRC psycholinguistic database: Machine-usable dictionary, version 2.00. Behavior Research Methods, Instruments, & Computers, 1988. 20(1): p. 6-10.
8. Vaden, K.I., H.R. Halpin, and G.S. Hickok, Irvine Phonotactic Onling Dictionary, Version 2.0, 2009.
9. Buz, E. and T.F. Jaeger. Word confusability and word durations. in IWoLP. 2012. New York.
10. Munson, B. and N.P. Solomon, The effect of phonological neighborhood density on vowel articulation. Journal of speech, language, and hearing research, 2004. 47: p. 1048-58.
11. Jaeger, T.F., The Ideal Speaker, 2011: University of Rochester.




**Supplemental figures**

The bar plots above show the residual word durations for a given contrast (ph1/ph2, where ph1 is the phone in the high confusability word that contrasts it with the low confusability word). Note that they can change from measure to measure.

The plot to the right shows how these word pairs look situated in the continuous version of the measure, in this case fNHD. The lines connect a given word pair's residual log duration means.