# Word Confusability and Word Durations

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

## Introduction

- **Communicative efficiency** accounts: contextually confusable words are pronounced with more effort (e.g. longer duration, less centered vowels) [1]
- **Production-centered** accounts have been claimed to predict the opposite [2-5], although mutually incompatible production accounts have been evoked [2,6]
- Previous work has used *out-of-context* measures of confusability (NHD, fNHD, etc.) [5,6-10], although comprehension is highly context-sensitive [11-14]
- For out-of-context (isolated) word production in past lab-studies, this might be acceptable. But recent work has begun to investigate in-context production in conversational speech [1,5,10]
- It is possible that the apparent conflict between these two lines of work is (at least partly) due to the failure to account for context (for preliminary evidence, see [15,16])

### Our Questions

1. Do speakers produce words that are contextually confusable differently from less confusable words?
2. Can results from lab-based and conversational speech be reconciled once context is taken into account?

## Study 1: Context in Corpus

**Goal**

Assess contextual confusability effects in conversational speech

**Hypothesis**

Higher bigram weighted NHD (CND) → longer duration

### Data

- Nouns, verbs and adjectives from the Switchboard Corpus
  - Removed types: <20 occurrences, >1 syllable, =0 phonological neighbors
  - Removed tokens: surrounding disfluency, speech duration or rate abs(z-score) > 2.5

### Analysis

- Model residual log duration with linear regression by word type
- Control measures used in residualization:
  - Expected duration & local speech rate
  - Log frequency, forward and backward bigram probability
  - Prior word mentions & distance (in words) since last mention
  - NHD and either Backward CND or Forward CND (given IV of interest)
  - Random by-speaker intercepts
- Investigate Forward and Backward CND (e.g., Forward CND($w_i$ | $w_{i-1}$) = $\Sigma_k$ p ($N_k(w_i)$ | $w_{i-1}$) / (1 – p($w_i$ | $w_{i-1}$), where $N_k(w_i)$ is the *k*th phonological neighbor of $w_i$)

### Results

- **Higher Forward CND → longer durations for all words** (β=0.009, 0.006, 0.004; t=7.6, 5.2, 1.7; nouns, verbs, adjectives; Fig 1.)
- **Higher Backward CND → longer durations for adjectives** (β=0.01; t=5.5) **but shorter durations for verbs** (β=-0.008; t=-7.5; Fig 2.).

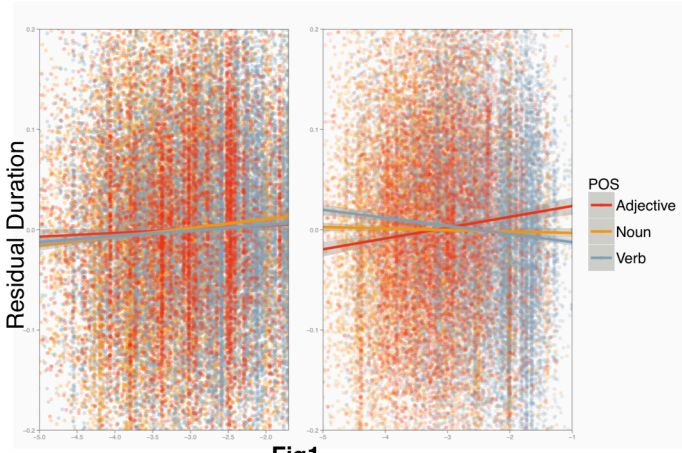



**Fig1.**

## Study 2: Context in Experiment

**Goal**

Same as Study 1 but in the lab

**Hypothesis**

Higher density after a neighbor → longer duration

### Data

- Speech from 36 (1 removed) undergrads naming 100 pictures
  - 20 pairs of monosyllabic neighbors (two removed), one high one low NHD (plus 60 fillers)
  - Neighbor pairs occur next to each other in experiment list (Fig 2)
  - Across subjects design, half of participants saw pair in one order
  - Hand correction of auto-alignment in progress (~8%)

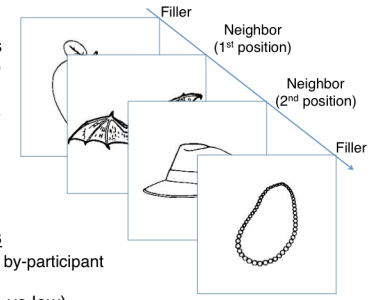



**Fig 2. Procedure**

### Analysis

- Model log word duration (random by-participant and by-trial intercepts)
- Compare effects of density (high vs low), position (first vs second) and density x position

### Results

- **High density → longer durations** (β=0.006; t=2.4; Fig 3.)
- **No effect of position or interaction**

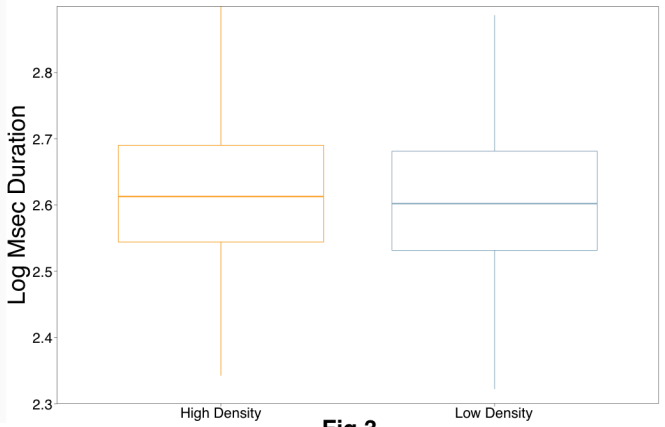
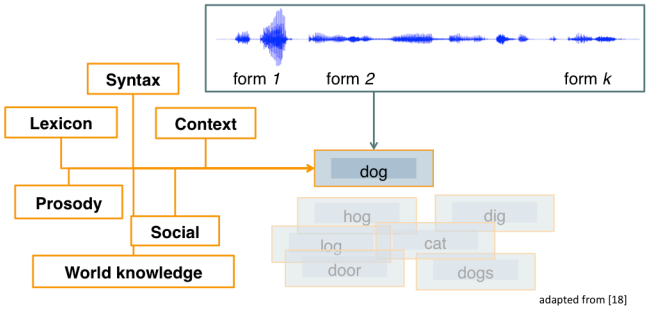



**Fig 3.**

## Conclusion

- **Conversational speech (corpus):**
  - Contextual NHD effects (except for backward CND for verbs) consistent with communicative efficiency (cf., 'ideal speaker model', [1]), replicating previous lab studies [15,16]
  - Replicating NHD effect of Gahl et al. (2012), while simultaneously showing why their arguments are problematic.
- **Picture description (lab):**
  - Duration, effects in line with other studies [7-9,15]
  - No context effects, but auto-aligned data (check back)
- **Moving forward:** mapping changes in production to intelligibility (cf., [3,17])
  - Communicative efficiency predicts greater *intelligibility*
  - Duration and dispersion as proxies for changes in intelligibility seems inappropriate given conflicting results

**Fig 4. Form may change based on *expected* confusability/competition**




References

1. Jaeger, T.F., 2011a.
2. Arnold, J.E., 2008.
3. Bard, E.G., et al., 2000.
4. Bell, A., et al., 2009.
5. Gahl, S., Y. Yao, and K. Johnson, 2012.
6. Baese-Berk, M. and M. Goldrick, 2009.
7. Munson, B., 2007.
8. Munson, B. and N.P. Solomon, 2004.
9. Scarborough, R., 2010.
10. Yao, Y., 2011.
11. Tanenhaus, M.K., et al., 1995.
12. Garnsey, S.M., et al., 1997.
13. McDonald, S.A. and R.C. Shillcock, 2003.
14. Trueswell, J.C., M.K. Tanenhaus, and C. Kello, 1993.
15. Heller, J., H. Lehnert-LeHouillier, and M. Goldrick, 2010.
16. Heller, J. and M. Goldrick, 2011.
17. Galatti, A. and S.E. Brennan, 2010.
18. Jaeger, T.F. 2011b.

Intended labels of stimuli

| High Density | Low Density |
|---|---|
| bark | shark |
| car | jar |
| bat | hat |
| man | pan |
| door | dog |
| fork* | cork* |
| mouse | house |
| crown | clown |
| net | pet |
| bear | hair |
| rain* | chain* |
| sock | rock |
| wig | pig |
| ring | king |
| bone | bowl |
| road | rope |
| heel | wheel |
| book | cook |
| sun | gun |
| sink | wink |

* Removed from analysis

**Lab experiment details:**

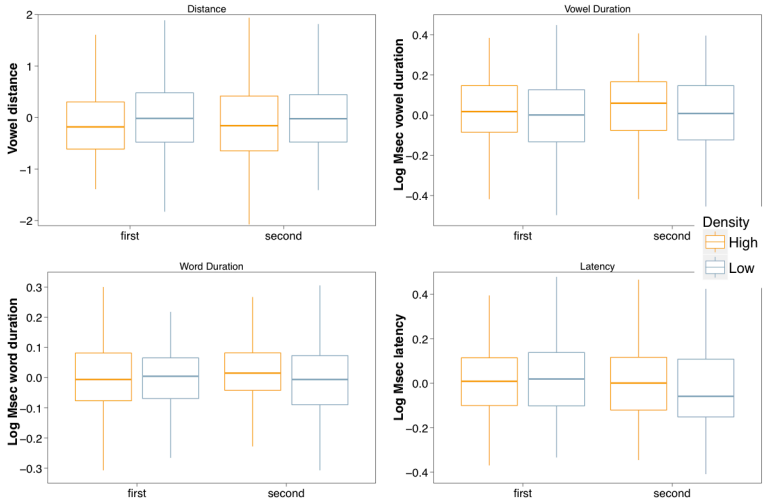
- Items were excluded if participants failed to label the pictures as intended. Two pairs were removed due to labeling accuracy 2.5 std dev away from mean. One subject removed due to labeling accuracy 2.5 std dev away from mean.
- F1 and F2 formant frequencies were measured at the vowel center then converted to the Bark scale. Vowel centers were estimated by subject across all labels given (intended/correct or not) except those that occurred after a disfluency or as part of a phrase.
- ~8% of data currently segmented by hand. The rest were generated using the Prosodylab-aligner (Gorman, Howell & Wagner, 2011)

**Additional analyses:**

Further analyses were done looking at vowel distance from center of vowel space, vowel duration, and picture naming latency. We present these here with the caveat that the data is still undergoing hand correction and that under an Ideal Speaker Model (Jaeger, 2011) we do not predict distance effects on the vowel given that vowels were shared across neighbor pairs (i.e. an Ideal Speaker would predict effects on the areas of neighbors that *differed*). Vowel distance and duration were different across the density groups. Vowel duration qualitatively mirrored word duration in effect and vowel distance showed the *opposite* of standard NHD effects namely the vowels in the high density group were centralized. Naming latency showed an interaction between density group and position, low density words in second position showed shorter latencies.

Lab Experiment Linear Model Summaries

| | Vowel Distance | Vowel Duration | Word Duration | Naming Latency |
|---|---|---|---|---|
| (Intercept) | −0.006 | −0.017 | −0.002 | 0.001 |
| | (0.110) | (0.029) | (0.014) | (0.019) |
| group = high density | −0.041* | 0.023*** | 0.006* | 0.001 |
| | (0.017) | (0.004) | (0.003) | (0.004) |
| position = second | | | | −0.016 |
| | | | | (0.010) |
| interaction | | | | 0.008* |
| | | | | (0.004) |
| Deviance | 1700.025 | −983.253 | −1923.265 | −1142.450 |
| BIC | 1743.245 | −934.585 | −1872.159 | −1062.414 |
| N | 967 | 967 | 967 | 967 |




**References**


Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes, 23*, 495-527.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and cognitive processes, 24*, 527-554.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M. P., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the Intelligibility of Referring Expressions in Dialogue. *Journal of Memory and Language, 42*, 1-22.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*, 92-111.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*.

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*, 35-51.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The Contributions of Verb Bias and Plausibility to the Comprehension of Temporarily Ambiguous Sentences. *Journal of Memory and Language, 93*, 58-93.

Heller, J., & Goldrick, M. (2011). *Context matters: effects of repetition and lexical neighborhood on vowel production.* Paper presented at the Testing Models of Phonetics and Phonology, Boulder, CO.

Heller, J., Lehnert-LeHouillier, H., & Goldrick, M. (2010). *The dynamics of lexical neighborhood effects on phonetic processing.* Paper presented at the 6th International Workshop on Language Production, Edinburgh, UK.

Jaeger, T. F. (2011). *The Ideal Speaker.* University of Rochester.

Jaeger, T. F. (2011, August 24th). *Linguistic Experience and Speakers' Contribution to Inference Over a Noisy Channel.* Paper presented at the Workshop on Structural Alternations: Speaker and Hearer Perspective, Groningen, Holland.

McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological science, 14*, 648-652.

Munson, B. (2007). Lexical Access , Lexical Representation , and Vowel Production. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 201-228): Mouton de Gruyter.

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of speech, language, and hearing research : JSLHR, 47*, 1048-1058.

Scarborough, R. (2010). Lexical and contextual predictability: Confluent effects on the production of vowels. In C. Fougeron, B. Kuehnert, M. Imperio & N. Vallee (Eds.), *Laboratory Phonology 10* (Vol. 10, pp. 557-586). Berlin; New York: De Gruyter Mouton.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. [Research Support, Non-U.S. Gov't
Research Support, U.S. Gov't, Non-P.H.S.
Research Support, U.S. Gov't, P.H.S.]. *Science, 268*(5217), 1632-1634.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 528-553.

Yao, Y. (2011). *The effects of phonological neighborhoods on pronunciation variation in coversational speech.* (PhD), University of California, Berkeley.