Effects of Phonological Confusability on Speech Duration

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

Introduction

- **Audience design** accounts predict that contextually confusable words are pronounced with more effort (e.g. longer duration). **Production-centered** accounts predict the opposite (e.g. Arnold, 2008; Bard et al., 2000; Bell et al., 2009; Gahl et al., 2012)
- Previous work (e.g. Munson, 2007; Munson and Solomon, 2004; Scarborough, 2010; Yao, 2011 Gahl et al, 2012) has exclusively used *out-of-context* measures of confusability, phonological neighborhood density (NHD) (Luce & Pisoni, 1998)
- But lab and natural corpus studies conflict in their findings
- Based on recent results (Heller et al., 2010; Heller and Goldrick, 2011), we hypothesize that the apparent conflict is due to the failure to account for context.

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: Gahl et al 2012 Replication

Goal
Test Gahl et al (2012) using the Switchboard Corpus
Hypothesis
High density nouns and verbs → shorter duration (replication)

Data Set

- Nouns and verbs extracted from the Switchboard Corpus
  - Removed types with: fewer than 20 occurrences, more than 7 or less than 3 phonemes, no phonological neighbors in Switchboard
  - Removed tokens with: pauses (filled or not) or disfluencies just prior or after, speech duration or speech rate absolute z-score > 2.5
- Final set: 94656 tokens (472 noun and 407 verb types)

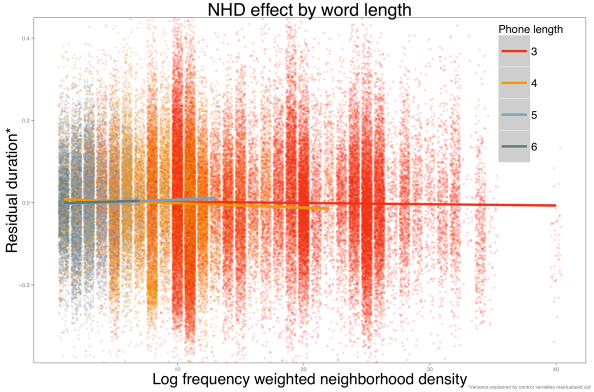
Analysis

- Model log duration with mixed effects linear regression
- Control measures:
  - Expected duration & speech rate
  - Log frequency, forward and backward bigram probability
  - Prior word mentions & distance (in words) since last mention
  - By speaker random intercepts
- Look for effects of log frequency weighted NHD by word length

Results

- **Higher NHD → shorter durations for 3 and 4 phoneme words** (β=-0.0003, -0.001; t=-3.8, -5.6), **the opposite was observed for 5 phoneme words** (β=0.001; t=2.7)

NHD effect by word length



Study 2: Context effects

Goal
Extend model from Study 1 with *contextual* confusability measures
Hypotheses
1) Higher bigram weighted NHD (CND) → longer duration
2) More neighbor mentions → longer duration
3) Shorter distance since last neighbor → longer duration

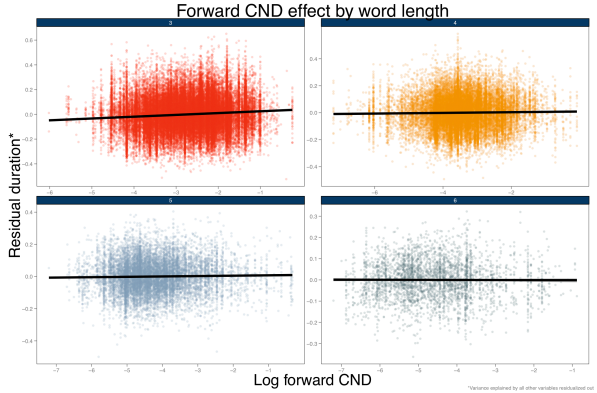
Analysis

- To models from Study 1 add:
  - Forward and backward CND (forward CND($w_i \mid w_{i-1}$) = Σ_k $p(N_k(w_i) \mid w_{i-1}) / (1 - p(w_i \mid w_{i-1})$, where $N_k(w_i)$ is the *k*th phonological neighbor of w_i)
  - Neighbor mentions
  - Distance since last mention

Results

- **Higher forward CND → longer durations for 3, 4 & 5 phoneme words** (β=0.025, 0.004, 0.004; t=27, 3.7, 3.1)
- **Higher backward CND → shorter durations for 3 & 6 phoneme words** (β=-0.03, -0.016; t=-36, -7.3) and the **opposite effect for 4 & 5 phoneme words** (β=0.011, 0.004; t=12, 2.8)
- Neighbor mentions not significant
- **Shorter distance since last neighbor → longer durations for 3, 4 & 5 phoneme words** (β=-0.00002, -0.00002, -0.00001; t=-5.6, -5.3, -2.6)

Forward CND effect by word length



Conclusion

General Discussion

- Replicate NHD effects from previous corpus study (e.g. Gahl et al 2012) but also find contextual NHD effects similar to those in previous lab studies (e.g. Scarborough, 2010; Heller & Goldrick, 2011)
- **Higher NHD may facilitate speech production**, though it remains unclear why facilitation results in reduction rather than clearer articulation (not addressed in the literature, cf. Arnold, 2008; Bard et al., 2000; Bell et al., 2009; Gahl et al., 2012 vs. Baese-Berk and Goldrick, 2009)
- CND, rather than NHD, is arguably more relevant for test of hypothesis that language production is organized for efficient communication (cf. 'ideal speaker model', Jaeger, 2011)
- **Results of CND expected if speakers strike efficient balance between production effort and intelligibility, but not by purely production-centered accounts.**

- **Moving forward** – how to provide an informative test of audience design hypothesis:
  1. Don't use out-of-context measures of confusability to test in-context production …
  2. Better models of context (e.g. integrating the various predictors employed here)
  3. Remove simplifying assumption shared in literature (e.g. word boundaries are *not* known)
  4. Measure intelligibility (cf. Bard et al., 2000; Galati and Brennan, 2010)

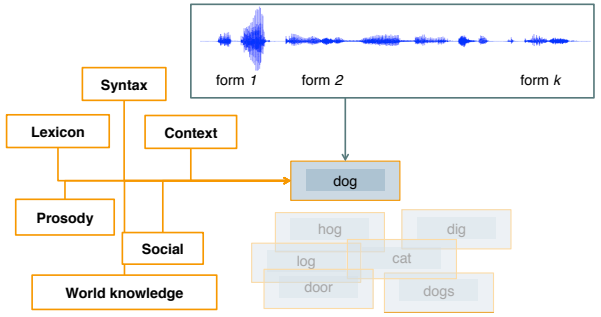
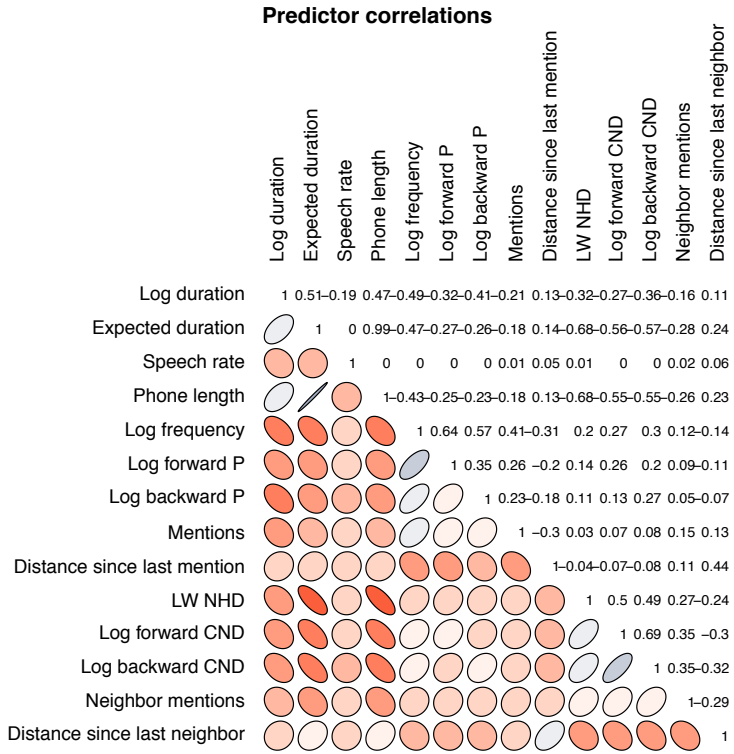


Table 1: Model outputs by phone length, Study 1

	3 Phones	4 Phones	5 Phones	6 Phones
(Intercept)	−0.000	−0.000	−0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.002)
Expected duration	0.158***	0.185***	0.167***	0.115***
	(0.003)	(0.004)	(0.004)	(0.007)
Speech Rate	−0.369***	−0.372***	−0.362***	−0.383***
	(0.007)	(0.009)	(0.011)	(0.020)
Log frequency	−0.051***	−0.010***	−0.030***	0.005
	(0.002)	(0.002)	(0.002)	(0.005)
Log forward P	−0.006***	−0.008***	−0.004***	−0.003
	(0.001)	(0.001)	(0.001)	(0.002)
Log backward P	−0.028***	−0.041***	−0.026***	−0.029***
	(0.001)	(0.001)	(0.001)	(0.002)
Mentions	−0.002***	0.000	0.000	−0.000
	(0.000)	(0.000)	(0.001)	(0.001)
Distance since last mention	−0.000*	−0.000***	−0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Log weighted NHD	−0.000***	−0.001***	0.001**	0.001
	(0.000)	(0.000)	(0.000)	(0.002)
Deviance	−60519.746	−37830.728	−19483.067	−6318.622
BIC	−60278.157	−37601.202	−19269.508	−6128.618
N	52761	26979	11556	3360

Table 2: Model outputs by phone length, Study 2

	3 Phones	4 Phones	5 Phones	6 Phones
(Intercept)	−0.000	−0.000	−0.001	−0.000
	(0.001)	(0.001)	(0.001)	(0.002)
Expected duration	0.151***	0.195***	0.170***	0.106***
	(0.003)	(0.004)	(0.004)	(0.007)
Speech Rate	−0.368***	−0.367***	−0.360***	−0.373***
	(0.007)	(0.009)	(0.011)	(0.020)
Log frequency	−0.051***	−0.011***	−0.026***	0.016**
	(0.002)	(0.002)	(0.003)	(0.005)
Log forward P	−0.008***	−0.009***	−0.006***	−0.002
	(0.001)	(0.001)	(0.001)	(0.002)
Log backward P	−0.023***	−0.042***	−0.026***	−0.025***
	(0.001)	(0.001)	(0.001)	(0.002)
Mentions	−0.001***	0.001*	0.002	−0.001
	(0.000)	(0.001)	(0.001)	(0.001)
Distance since last mention	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Log weighted NHD	−0.000**	−0.003***	0.000	0.004**
	(0.000)	(0.000)	(0.000)	(0.002)
Log forward CND	0.025***	0.004***	0.004**	−0.001
	(0.001)	(0.001)	(0.001)	(0.002)
Log Backward CND	−0.030***	0.011***	0.004**	−0.016***
	(0.001)	(0.001)	(0.001)	(0.002)
Neighbor mentions	−0.000	0.001	−0.002	0.001
	(0.000)	(0.001)	(0.002)	(0.005)
Distance since last neighbor	−0.000***	−0.000***	−0.000**	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Deviance	−61966.915	−38225.100	−19530.366	−6420.258
BIC	−61618.167	−37894.155	−19223.024	−6146.080
N	52761	26979	11556	3360



References

1. Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes, 23*, 495-527.
2. Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and cognitive processes, 24*, 527-554.
3. 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.
4. 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
5. Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*.
6. 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.
7. 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.
8. 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.
9. Jaeger, T. F. (2011). *The Ideal Speaker*. University of Rochester.
10. Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear and hearing, 19*, 1-36.
11. Munson, B. (2007). Lexical Access , Lexical Representation , and Vowel Production. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 201-228). New York: Mouton de Gruyter.
12. 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.
13. Scarborough, R. (2010). Lexical and contextual predictability: Confluent effects on the production of vowels. In C. Fougeron, B. Kuehnert, M. Imperio & N.