

Effects of phonological confusability on speech duration

Esteban Buz & T. Florian Jaeger (University of Rochester)

ebuz@bcs.rochester.edu

Speech production; Natural corpora; Confusability

When speaking, we need to balance the wish to be understood against bottlenecks resulting from the planning processes involved in language production. How speakers manage this balance is one of the central questions in psycholinguistics. Some posit that speakers merely hem to their internal production constraints while others posit that speakers strike a balance between their internal constraints and those of their listeners (audience design). Work in the intelligibility of speech has focused on if and how speakers modify their articulations so as to avoid confusion. Currently, there are two opposing findings in confusability and articulation: in scripted sentence and isolated word production there is evidence that phonological confusability (as approximated by phonological neighborhood density, NHD) results in greater vowel dispersion and longer spoken duration [1]. Conversely, in unscripted speech the conclusion has been that NHD results in shorter durations and less vowel dispersion [2] & [3]. Recent work suggests that articulatory efforts are a function of *contextual* confusability [4].

We hypothesize that language production is sensitive to *contextual* confusability and that this explains the apparent conflicts in prior work. We analyzed word durations in Switchboard as a function of three novel measures of contextualized confusability: bigram-weighted NHD (CND, i.e. $\text{forward CND}(w_i | w_{i-1}) = \sum_k p(N_k(w_i) | w_{i-1}) / (1 - p(w_i | w_{i-1}))$, where $N_k(w_i)$ is the k th phonological neighbor of w_i), number of prior neighbor mentions and distance since last neighbor.

Data: We extracted all nouns and verbs from Switchboard[5]. We excluded tokens preceding or following a speech pause or disfluency, types with fewer than 20 occurrences, types with more than 7 phonemes, tokens with missing data and tokens with absolute log duration and log speech rate z-scores > 2.5. The final set contained 99053 tokens (542 noun and 433 verb types).

Analysis: Normal NHD and our novel measures of confusability were included in a mixed effect linear regression against log-transformed duration while controlling for expected word duration based on biphone duration averages, phoneme length, contextual log syllable per second speech rate, log frequency (in Switchboard), bigram forward and backward probability, distance in words since last mention and random by-speaker intercepts.

Results: As in [2] & [3], greater NHD reflected shorter durations ($ps < .05$), this suggests that speakers do not adjust for context neutral confusability as measured by NHD. Consistent with [4], as distance since last neighbor increased, spoken durations decreased ($ps < .05$); further, higher forward CND resulted in longer durations ($ps < .05$). Backward CND had a mixed effect on duration; more work is needed to know why.

Conclusion: While we replicate the finding that greater out-of-context NHD seems to be correlated with shorter durations (contrary to the audience design hypotheses), we also find that word durations tend to be longer if a word would (otherwise) be confusable *in its actual context* based on CND and proximity to neighbors (consistent with the audience design hypotheses). More broadly this work suggests divergence between unscripted and scripted speech patterns with regards to NHD is partially explained by contextual factors.

References

1. Scarborough, R., *Lexical and contextual predictability: Confluent effects on the production of vowels*, in *Laboratory Phonology 10*, C. Fougerson, et al., Editors. 2010, De Gruyter Mouton: Berlin; New York. p. 557-586.
2. Yao, Y., *The effects of phonological neighborhoods on pronunciation variation in conversational speech*, in *Department of Linguistics 2011*, University of California, Berkeley.
3. Gahl, S., Y. Yao, and K. Johnson, *Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech*. *Journal of Memory and Language*, 2012.
4. Heller, J. and M. Goldrick. *Context matters: effects of repetition and lexical neighborhood on vowel production*. in *Testing Models of Phonetics and Phonology*. 2011. Boulder, CO.
5. Calhoun, S., et al., *The NXT-format Switchboard Corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue*. *Language Resources and Evaluation*, 2010. 44: p. 387-419.