Phone Information Content Influences Phone Duration

Uriel Cohen-Priva and Dan Jurafsky

Department of Linguistics, Stanford University, CA 94305, USA
{urielc,jurafsky}@stanford.edu

Recent years have seen a growing interest in using information theoretic terms to explain linguistic phenomena (Aylett and Turk, 2004; Levy and Jaeger, 2006; Raymond et al., 2006), as well as a growing availability of large scale phonetic transcription of spontaneous speech (Greenberg et al., 1996; Pitt et al., 2007). Previous work has shown, for example, that predictable function words are shorter (Jurafsky et al., 2001), that syllables with predictable phone sequences are shorter (Aylett and Turk, 2004), and that morphemes are shorter when they are predictable from surrounding words (Pluymaekers et al., 2005).

We propose to study a new source of information: *phone informativity*, the amount of information contained in each phone. We propose that phones differ in the amount of information they contain, and that this information value influences the surface realization of the phone: uninformative phones are shorter than informative phones. This idea explains interesting asymmetries that are difficult to account for using phonological variables: /k/ is less informative and shorter than /p/, while /g/ is more informative and longer than /b/. Moreover, it goes beyond predictabity, as it explains why the highly uninformative /ŋ/ is shorter than the highly informative /m/, even when /ŋ/ is unpredictable, and /m/ is.

In order to compute phone informativity, we first operationalize the basic information value of a phone in context as its contribution to word recognition: the negative log probability of seeing a phone given all the previous phones in the word, $-\log_2(\Pr(\text{phone}|\text{previous phones}))$. This means that the more likely we are to see a phone in a context, the less informative it is in that context. We then estimate the *phone informativity* as the mean information value of all its occurrences.

We tested this claim by building a model to predict the duration of consonants in the Buckeye corpus of phonetically transcribed speech (Pitt et al., 2007). We used a linear regression to control for the phonological properties of the segments, phonological properties of neighboring segments, stress, rate of speech, and previous information theoretic variables used in (Aylett and Turk, 2004) such as word frequency and phone bigram and trigram probability to predict the duration of the consonants. After controlling for these factors, the phone informativity was a strong predictor of duration: more informative phones were longer. Indeed, the phone informativity was a stronger factor than all other information-theoretic factors such as word frequency and contextual biphone and triphone phone predictability (see Table 1).

Our results are the first to confirm the important effect of earlier-studied variables like word frequency and biphone predictability on phone duration (where previous studies had only looked at syllable or morpheme duration). Our results confirm the important role of phone informativity in phonetic production, suggesting that information value is a key part of the representation of phones, and influences the effort made by speakers to convey meaning.

Table 1: A 'drop1' (Chambers, 1992) Analysis of Select Variables

Variable	Df	Sum of Squares
Rate of speech	1	522
The phone is ±VOICE	1	515
The PLACE of the phone	3	602
The next phone is ±VOICE	2	156
Word unigram probability	1	72
Phone's bigram probability	1	69
Phone informativity	1	**202**

References

Aylett, M. and Turk, A. (2004). Explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47.

Chambers, J. M. (1992). *Linear models*, chapter 4. Wadsworth and Brooks / Cole.

Greenberg, S., Ellis, D., and Hollenback, J. (1996). Insights into spoken language gleaned from phonetic transcription of the Switchboard corpus. In *icslp96*, pages S24–27, Philadelphia, PA.

Jurafsky, D., Bell, A., Gregory, M. L., and Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J. L. and Hopper, P., editors, *Frequency and the Emergence of Linguistic Structure*, pages 229–254. Benjamins, Amsterdam.

Levy, R. and Jaeger, T. F. (2006). Speakers optimize information density through syntactic reduction. In *proceedings of the Twentieth Annual Conference on Neural Information Processing Systems*.

Pitt, M., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., and Fosler-Lussier, E. (2007). Buckeye corpus of conversational speech (2nd release). Department of Psychology, Ohio State University.

Pluymaekers, M., Ernestut, M., and Baayen, R. H. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, 62:146–159.

Raymond, W. D., Dautricourt, R., and Elizabeth, H. (2006). Word-medial /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, 18.