Phone Information Content Influences Phone Duration

Uriel Cohen-Priva, Dan Jurafsky, Department of Linguistics, Stanford University

Conference on Prosody and Language Processing. April 11-13 2008 at Cornell University

Abstract

We show that the information value of a phone influences its duration. Informative phones like /g/ are longer than uninformative phones like /d/. This extends previous work showing that predictability affects word and morpheme duration. Uninformative phones are short even when they appear in informative (unpredictable) contexts.

Introduction

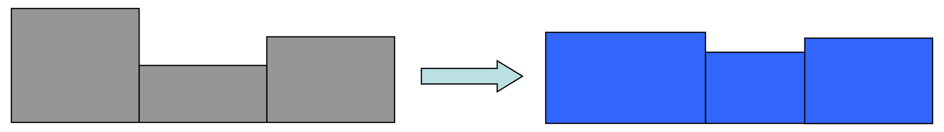
Background

Uniform Information Density/Smooth Signal Hypothesis

❖ Contextually predictable words and their subparts (morphemes, syllables, segments) are shorter than unexpected words.

❖ This may play a role in smoothing information across the signal, mediated by prosody.

(Schuchardt 1885, Zipf 1929, Fidelholz 1975, Hopper 1976, Aylett 1999, Jurafsky et al. 2001, Pluymaekers et al. 2005, van Son and van Santen 2005, Aylett and Turk 2004, 2006, inter alia)



Segment Simplicity

❖ Underspecification theories assume that some segments are simpler (specify fewer values) than other phones. *(Kiparsky 1993, inter alia)*

❖ Markedness accounts claim there is a hierarchy of simplicity that is inherent to human language. *(de Lacy 2002, inter alia)*

Present Study

❖ Are certain phones simpler information theoretically?

❖ Are simpler (uninformative) phones shorter?

➢ In addition to (known) effect of contextual predictability

❖ Does this correlate with underspecification or markedness?

Method

Procedure

Test the effect of ***phone informativity*** on phone duration.

Phone Informativity

❖ We estimate a phone's ***surprisal*** as the log probability of seeing a phone given previous phones in the word.

log P(phone | previous phones in word)

❖ We estimate a phone's ***informativity*** to be the mean **surprisal**, across the words in the lexicon, with each word weighted by its token frequency.

Corpora and Data

❖ Token frequencies estimated from the Switchboard and Buckeye corpora.

❖ Pronunciations are from the CMU dictionary *(CMU 1993)*.

❖ Phone durations are from the Buckeye corpus of phonetic transcriptions *(Pitt et al 2005, 2007)*

❖ We limited ourselves to word-medial, non-deleted consonants.

❖ Onsets and codas were estimated separately.

Controls

❖ The phonological attributes of the phone and adjacent phones

❖ Syllable stress

❖ Word and phone frequency

❖ Local predictability, based on previous one / two phones (bi / tri-phones)

❖ Rate of speech

Method

Estimation Method

We used linear regression to estimate the log phone duration.

❖ The duration of a phone is estimated to be a product of several factors.

❖ The linear regression estimates the weight each factor has.

➢ Phone duration is λ times longer for each value a categorical factor has, and λ is estimated by the regression.

➢ Phone duration is raised to the power of real-valued parameters, weighted by some λ.

❖ We check how significant each coefficient is, and in what direction it influences duration.

❖ Significance is estimated by comparing the residual sum of squares of two almost-identical models, one containing the variable in question, and one that does not contain it.

Results

Phone informativity is a very strong predictor of phone duration.

❖ For all phones, the more informative they are, the longer they are.

❖ Onset phones are longer when the previous phone is more informative.

❖ Coda phones are longer when the following phone is less informative.

❖ These effects are stronger than local predictability effects (bi-phone / tri-phone).

A Comparison of Factors (codas)	Sum of Squares / Df	p	Duration
Log phone probability, given previous phone (the same estimate using two phones was not significant)	1.1 / 1	0.03573	Longer
Log phone probability, ignoring context	1.8 / 1	< 0.01	Longer
Phone informativity	11 / 1	$< 10^{-10}$	Longer
Next phone informativity	7.1 / 1	$< 10^{-7}$	Shorter

A Comparison of Factors (onsets)	Sum of Squares / Df	p	Duration
Log phone probability, given one or two previous phones	9.2 / 2	$< 10^{-11}$	Longer
Log phone probability, ignoring context	47.7 / 1	$< 10^{-16}$	Longer
Phone informativity	21.4 / 1	$< 10^{-16}$	Longer
Previous phone informativity	15.5 / 1	$< 10^{-16}$	Longer

For each class of stops, *duration matches informativity*

❖ /p/ >> /k/ >> /t/

❖ /g/ >> /b/ >> /d/

❖ /m/ >> /n/ >> /ŋ/

Voiceless stops	/p/	/t/	/k/
Informativity	3.93	1.61	2.75
Mean duration	0.082	0.055	0.077

Voiced stops	/b/	/d/	/g/
Informativity	3.96	1.90	4.8
Mean duration	0.058	0.039	0.060

Nasal stops	/m/	/n/	/ŋ/
Informativity	3.07	1.8	0.24
Mean duration	0.061	0.057	0.048

Conclusion

❖ **Phone informativity has a strong influence on phone duration.**

❖ **Consistent with the smooth signal hypothesis.**

➢ Lengthening a more informative phone provides a more even distribution of information.

➢ Lengthening a phone after an informative phone allows spillover.

➢ Problem: Why shorten a phone when a highly informative phone follows?

❖ **Partly consistent with markedness or underspecification.**

➢ Coronals are indeed less informative, underspecified and unmarked.

➢ But not clear why dorsals would be more marked than labials, or the other way around.

❖ **The mental representation of phones includes their informativity.**

➢ A contextually unpredictable phone with low informativity may be shorter than a contextually predictable but highly informative phone.

- This is true even after we control for phone frequency. *(Zipf 1929)*
- This is true even after we control for local predictability factors. *(Aylett and Turk 2004, inter alia)*

Future Directions

❖ **Cohen-Priva (2008) has similar results for medial phone deletions.**

❖ **Vowel duration models are trickier:**

➢ The base measurement, *surprisal*, has to control for stress.

➢ Vowels are greatly affected by local context.

➢ Informativity shows in duration stability: informative vowels vary less.

❖ **Promising but incomplete results for reduction models.**

❖ **How much of historical change is derived by informativity?**

➢ Use informativity to explain why different dialects of English delete and change /t/, /d/, /r/, /n/ and /ŋ/, which have extremely low informativity.

➢ In Latinate languages, similar processes target /s/.

References

Aylett, M. P. (1999). Stochastic suprasegmentals: Relationships between redundancy, prosodic structure and syllabic duration. *Proceedings of the 1999 International Conference on Spoken Language Processing* (pp. 289-292).
Aylett, M., and Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, **47**, 31–56.
Aylett, M., and Turk, A. (2006). Language redundancy predicts syllable duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, **119**, 3048-3058.
Cohen-Priva, U. (2008). *Using information content to predict phone deletion*. Talk to be given at WCCFL 27, UCLA.
CMU. 1993. The Carnegie Mellon Pronouncing Dictionary v 0.1. Carnegie Mellon University
Fidelholz, J. (1975). Word frequency and vowel reduction in English. *CLS-75* (pp. 200–213). Chicago: University of Chicago.
Greenberg, S., Ellis, D., and Hollenback, J. (1996). Insights into spoken language gleaned from phonetic transcription of the Switchboard corpus. In *ICSLP 96*, pages S24–27, Philadelphia, PA.
Hooper, J. B. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In Christie, W. (Ed.), *Current progress in historical linguistics* (pp. 96–105). Amsterdam: North Holland.
Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J., and Hopper, P. (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam: Benjamins.
Kiparsky, P. 1993. Blocking in non-derived environments. In S. Hargus and E. Kaisse (Eds.), *Phonetics and Phonology 4: Studies in Lexical Phonology*. San Diego: Academic Press.
de Lacy, P. (2002). *The formal expression of markedness*. PhD thesis, University of Massachusetts Amherst.
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.
Pitt, M., Johnson, K., Hume, E., Kiesling, S. and Raymond, W. (2005). The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication*, **45**, 90-95.
Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, **62**, 146–159.
Schuchardt, H. (1885). *Über die Lautgesetze: Gegen die Junggrammatiker*. Berlin: Robert Oppenheim. Excerpted with English translation in Vennemann, T. and Wilbur, T. H. (Eds.). (1972). *Schuchardt, the Neogrammarians, and the transformational theory of phonological change* (pp.1-72). Frankfurt: Athenaum Verlag.
van Son, R.J.J.H. and van Santen, J.P.H. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication, *Speech Communication,* **47**, 100-123.
Zipf, G. K. (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology*, **15**, 1-95.

Acknowledgments

❖ We would like to thank Paul Kiparsky, Chris Manning, Arto Anttila, Matthew Adams and Nola Stephens for all their useful advice and comments.

❖ This research was supported by the NSF via Award #IIS-0624345