DURATIONAL VARIATION IN POLISH FRICATIVES PROVIDES EVIDENCE FOR HYBRID MODELS OF PHONOLOGY

Kamil Kaźmierski

Faculty of English at Adam Mickiewicz University, Poznań
kamil.kazmierski@wa.amu.edu.pl

ABSTRACT

The neighborhood density of a word is the number of words that sound similar to it. Phonotactic probability is a measure of how typical (for a given language) the phoneme sequences in a word are. These two factors are known to affect speech perception in opposing directions: high neighborhood density slows down processing while high phonotactic probability speeds it up [30]. This finding supports hybrid models of phonological representation [24], as neighborhood density effects operate on lexical, and phonotactic probability effects on sublexical representations.

The present paper, investigating word-initial double clusters retrieved from the Greater Poland Spoken Corpus [14], tests the predictions for durational variation in fricatives following from Vitevitch and Luce [30]. It has been found that high neighborhood density is associated with longer - while high phonotactic probability with shorter - fricative durations. Thus, further support for hybrid models of phonological storage is provided.

Keywords: hybrid models, mental lexicon, neighborhood density, phonotactic probability, corpus phonology

1. INTRODUCTION

There are two competing approaches to phonological representation: the abstractionist approach and the rich storage approach. The abstractionist approach is adopted in most of generative work and embraces the model of speech production proposed by Levelt et al. [17], with abstract lexical representations and feed-forward architecture of speech production involving discrete modules. The rich storage approach, associated with the exemplar theory [12, 4, 23], assumes phonetically rich representations and sees abstract lexical representations as, at best, epiphenomena. Recognizing the need for both levels of representation in the mental lexicon, so-called hybrid models of phonology have been proposed [24, 9, 5].

One of the experimental findings supporting hybrid models is presented by Vitevitch and Luce [30], who found divergent effects of phonotactic probability and neighborhood density on speech recognition. High phonotactic probability, they found, facilitates word recognition, while high neighborhood density hampers it. An interpretation of this finding that they propose is that the effect of phonotactic probability stems from the role of sublexical representations and the neighborhood density effect stems from the role of lexical representations in speech processing. As neighborhood density and phonotactic probability are positively correlated [16], Vitevitch and Luce [30] compared processing speed of words and nonce words. Their reason for using nonce words was that no lexical effects were to be expected for them, enabling the facilitatory effect of phonotactic probability to transpire. For the real words, conversely, the inhibitory effect of neighborhood density was expected to prevail, so that high-probability, high-density words were predicted to show longer reaction times than low-probability low-density words. This was indeed the case.

It has been argued that the structure of the mental lexicon can be manifested in the speech signal itself [1], and different types of phonological representations have been claimed to manifest themselves in speech production data [10]. The effects of neighborhood density and of phonotactic probability have indeed been attested in speech production. Low neighborhood density has been shown to be associated with more centralization in vowel production [22, 33], (though cf. [8]), and high neighborhood density has been found to be associated with higher degrees of coarticulation [27]. High phonotactic probability, on the other hand, is associated with greater accuracy in nonword repetition [31].

With multiple regression modeling it is now possible to investigate the influence of (weakly) correlated predictors [32]. Seizing on this opportunity, the present study investigates the divergent effects of neighborhood density and phonotactic probability in acoustic speech production data, drawn from a corpus [14] of unscripted speech. By definition, the study is limited to real words only. Using unscripted speech increases the ecological validity of results [29], by going beyond careful, laboratory speech style. This analysis has sought to ascertain the divergent effects of neighborhood density and phonotactic probability on the durational variation in fricatives. Fricatives in words with high neighborhood densities are predicted to show less durational reduction, i.e. longer durations, than fricatives in words with low neighborhood densities. This is in keeping with Lindblom's [18] model, according to which lexical items which might pose difficulty for the listener show less reduction than items posing less difficulty. As items with higher numbers of neighbors have been shown to slow down retrieval [19], their durations are expected to be longer. At the same time, fricatives in words of high phonotactic probability are predicted to show more durational reduction, i.e. shorter durations, than fricatives in words of low phonotactic probability. Higher levels of activation of frequent sublexical units should contribute to their greater temporal reduction.

2. METHOD

2.1. Data

Words (*N* = 2788, 702 distinct word-form types) beginning with double consonant clusters (one of the fricatives: /ɕ ʂ ʑ f s v x z/ followed by a plosive, nasal, liquid or glide) were retrieved from the Greater Poland Spoken Corpus [14]. The corpus contains interview speech of 63 speakers (50 female, 13 male), and includes phoneme-level and word-level annotations force-aligned using acoustic models trained with HTK [34] for each speaker individually in LaBB-CAT [7], an Open Source corpus management and annotation suite. Fricatives were chosen on the assumption that their segmentation would be relatively reliable. Since they are obstruents, their aperiodic noise makes their segmentation from preceding vowels easier than would be the case for sonorants. Their relatively stable acoustic cues are present throughout most of the production of fricatives. This makes them preferable to stops. The closure stage of a stop may be silent, and so its beginning is unidentifiable if following a pause, and acoustic cues to stop release often spill onto the following vowel. Clusters rather than singletons were decided on as words containing clusters provide the opportunity for a more fine-grained measurement of phonotactic probability.

2.2. Analysis

A mixed-effects linear regression model was fitted to the data with the *lme4* package [3] in R [26]. The use of a mixed-effects model, i.e. one involving random terms (intercepts and slopes) was called for since the observations are not independent: there are several observations for each speaker and for each item (cf. Baayen et al. [2]).

2.2.1 Outcome variable: Phonetic duration

The duration of the fricative, which was always the first element of the cluster, was log-transformed, centered, and entered as a continuous response variable. The log transformation, while making the interpretation less intuitive, was required since entering a raw duration measure resulted in the heteroskedasticity of residuals in an initially fitted model.

2.2.2 Test variables: Neighborhood density and phonotactic probability

The two main predictor variables for this study were phonotactic probability and neighborhood density. The two metrics were calculated using Phonological CorpusTools [11]. For the calculation of phonotactic probability, the algorithm developed by Vitevitch and Luce [30], based on average bigram positional probabilities across a word, was used. For neighborhood density, the standard approach of measuring string similarity (the Levenshtein edit distance) was used, i.e. the number of lexemes that would be formed by removing, substituting or adding one phoneme to a given word was computed. Both measures were calculated based on the transcriptions in the Greater Poland Spoken Corpus [14].

2.2.3 Co-variates

A number of covariates were included in the model to account for the variance stemming from factors known to influence durational variation. These were: average per speaker speech rate (a syllables per second measure of all speaker utterances in the corpus, a numerical predictor), speech rate deviation (the difference between the speech rate of a given utterance and a speaker's average rate, cf. [28], a numerical predictor), speaker gender (a binary predictor), word-form frequency (based on SUBTLEX-PL [21], numerical predictor), word duration (numerical predictor), morphological status (whether the fricative is a prefix, a binary predictor, cf. [25]), and stress (whether the initial syllable is stressed or not, cf. [20], binary predictor).

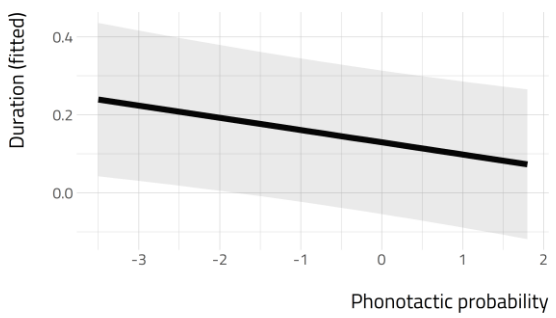
2.2.4 Random effect structure

To account for durational properties inherent to a given fricative [13, 15], a by-fricative random intercept term was included. To account for speaker-specific variation, the model included a by-speaker random intercept, as well as by-speaker random slopes for each of the test variables – neighborhood density and phonotactic probability – to allow for differences in the size of the effects of these variables on individual speakers.

3. RESULTS

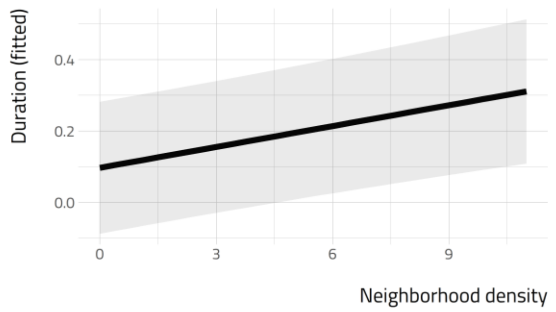
The model converged with Marginal R^2 of 0.106, and Conditional R^2 of 0.423. The discrepancy between the two measures attests to the importance of including random terms – a large portion of the variance that the model accounts for is taken care of by the random effects. The random effect affecting the duration of the fricative the most was the identity of the fricative (*SD* = 0.26), followed by the identity of the speaker (*SD* = 0.07). The by-speaker phonotactic probability random slope had a standard deviation of 0.04, and the by-speaker neighborhood density random slope had a standard deviation of 0.01. The inspection of a histogram and a Q-Q plot of the residuals (not shown) gave no reason to suspect the violation of the normality assumption.

Figure 1: An **increase** in phonotactic probability is associated with a **decrease** in fricative duration.



Moving on to fixed-effects, the results of modeling attest to a statistically significant association between a number of predictor variables and fricative duration. Fricatives produced at a speaking rate faster than usual for a given speaker are shorter ($\hat{\beta}$ = -0.05, $p <$ 0.001). Fricatives produced by speakers who generally speak fast relative to other speakers are also shorter ($\hat{\beta}$ = -0.05, $p < 0.001$). Somewhat more surprisingly, the longer the word, the longer the duration of the initial fricative ($\hat{\beta}$ = 0.93, $p < 0.001$). If the initial fricative forms a prefix, its duration is shorter ($\hat{\beta}$ = -0.12, $p < 0.001$) (agreeing with Plag et al.'s [25] results for the unvoiced fricative). If the initial fricative is in the onset of a stressed syllable, it is longer ($\hat{\beta}$ = 0.08, $p < 0.001$) (as in [20]). The effects of gender and lexical frequency have not reached statistical significance. A summary of all fixed effects is presented in Table 1.

Figure 2: An **increase** in neighborhood density is associated with an **increase** in fricative duration.



Crucially for the research hypothesis, larger phonotactic probability values are associated with shorter durations ($\hat{\beta}$ = -0.03, p = 0.011), while larger neighborhood density values are associated with longer durations ($\hat{\beta}$ = 0.02, $p < 0.001$). Partial effect plots in Figures 1 and 2 illustrate these two effects (the predictor effects were calculated by averaging over all other fixed effects, using the *effects* package [6]). The grey bands are 95% confidence intervals.

Table 1: Summary of fixed-effect coefficients in the linear regression model of durational variation: Coefficient estimates, standard errors, *z*, and corresponding *p*-value.

Predictor	$\hat{\beta}$	**SE ($\hat{\beta}$)**	***z*-value**	**Pr (>|*z*|)**
(Intercept)	-0.28	0.10	-2.80	**0.021**
Phonotactic probability	-0.03	0.01	-2.58	**0.011**
Neighborhood density	0.02	0.00	4.35	**<0.001**
Rate deviation	-0.05	0.01	-4.77	**<0.001**
Average rate	-0.05	0.01	-3.81	**<0.001**
Gender Male	-0.02	0.03	-0.59	0.557
Word duration	0.93	0.06	14.99	**<0.001**
Frequency	0.01	0.01	0.57	0.567
Prefix TRUE	-0.12	0.02	-4.80	**<0.001**
Stress TRUE	0.08	0.02	3.85	**<0.001**

4. DISCUSSION

The predictions following from the research hypothesis were confirmed. Within the same set of acoustic speech production data, statistically significant divergent effects of neighborhood density and phonotactic probability were found.

Neighborhood density effects, as they show the influence of the number of similar words on a given word's processing, point to an abstract, lexical level of processing. Lexical neighbors are related, or similar, with regard to abstract phonological entities. Phonotactic probability effects, on the other hand, as they show the influence of the frequency of occurrence of sublexical, positionally defined units on a given word's processing, point to a sublexical level of processing.

Consequently, further support is provided for hybrid models of phonology, such as [23, 9, 5]. Both lexical and sub-lexical levels of phonological processing, in production as well as in perception, need to be incorporated into phonological models to get a full picture of phonological storage in the mental lexicon.

These findings complement those of Vitevitch and Luce [30], who found divergent effects of neighborhood density and phonotactic probability in speech perception. Additionally, besides going from perception to production, this study has also tested the predictions following from the original study to another language, Polish, thus further extending the generalizability of the effect. Finally, a case for the usefulness of using unscripted speech corpus data, coupled with mixed-effects regression modeling for probing phonological questions has been made.

5. ACKNOWLEDGMENTS

This research was supported by the National Science Centre (Poland) grant No. UMO-2017/26/D/HS2/00027.

6. REFERENCES

[1] Baayen, R. H. 2007. Storage and computation in the mental lexicon. In: Jarema, G., Libben, G. (eds), *The mental lexicon: Core perspectives.* Oxford: Elsevier, 81–104.

[2] Baayen, R. H., Davidson, D. J., Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4), 390-412.

[3] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1-48.

[4] Bybee, J. 2001. *Phonology and language use*. Cambridge: CUP.

[5] Ernestus, M. 2014. Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua* 142, 27–41.

[6] Fox, J., Hong, J. 2009. Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package. *Journal of Statistical Software* 32(1), 1–24.

[7] Fromont, R. Hay, J. 2012. LaBB-CAT: An Annotation store. *Proc. of the Australasian Language Technology Association Workshop* Dunedin, New Zealand, 113–117.

[8] Gahl, S. 2015. Lexical competition in vowel articulation revisited: Vowel dispersion in the Easy/Hard database. *Journal of Phonetics* 49, 96–116.

[9] Goldinger, S. D. 2007. A complementary-systems approach to abstract and episodic speech perception. In: Trouvain, J., Barry, W. J. (eds), *Proc. 16th ICPhS* Saarbrücken, 49–54.

[10] Goldrick, M., Rapp, B. 2007. Lexical and post-lexical phonological representations in spoken production. *Cognition* 102, 219–260.

[11] Hall, K. C., Allen, B. Fry, M., Mackie, S., McAuliffe, M. 2018. Phonological CorpusTools, Version 1.3. [Computer program]

[12] Johnson, K. 1997. Speech perception without speaker normalization: An exemplar model. In: Johnson, K., Mullennix, J. W. (eds), *Talker variability in speech processing*. San Diego: Academic Press, 145–165.

[13] Jongman, A. 1989. Duration of frication noise required for identification of English fricatives. *The Journal of the Acoustical Society of America* 85 (4), 1718–25.

[14] Kaźmierski, K., Kul, M., Zydorowicz, P. In press. Educated Poznań speech 30 years later. *Studia Linguistica Universitatis Iagellonicae Cracoviensis*.

[15] Klessa, K. 2012. Polish segmental duration: selected observations based on corpus data. *Speech and Language Technology* 1112(1415), 94–104.

[16] Landauer T. K., Streeter L. A. 1973. Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior* 12(2), 119–131.

[17] Levelt, W. J. M., Roelofs, A., Meyer, A. S. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22, 1–75.

[18] Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H&H theory. In: Hardcastle, W., Marchal, A. (eds), *Speech production and speech modelling*. Dordrecht: Kluwer, 403–439.

[19] Luce, P. A., Pisoni, D. B. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19, 1–36.

[20] Łukaszewicz, B. 2018. Phonetic evidence for an iterative stress system: The issue of consonantal rhythm. *Phonology* 35, 115–150.
[21] Mandera, P. Keuleers, E. Wodniecka, Z., Brysbaert, M. 2015. SUBTLEX-PL: Subtitle-based word frequency estimates for Polish. *Behavior Research Methods* 47(2), 471–483.
[22] Munson, B.. Solomon, N. P. 2004. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47(5), 1048–1058.
[23] Pierrehumbert, J. B. 2002. Word-specific phonetics. In: Gussenhoven, C., Warner, N. (eds), *Laboratory Phonology VII*. Berlin: Mouton de Gruyter, 101–139.
[24] Pierrehumbert, J. B. 2006. The next toolkit. *Journal of Phonetics* 34, 516–530.
[25] Plag, I. Homann, J. Kunter, G. 2017. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* 53, 181–216.
[26] R Core Team. 2018. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
[27] Scarborough, R. A. 2004. Coarticulation and the structure of the lexicon. UCLA dissertation.
[28] Tanner J., Sonderegger, M., Wagner, M. 2017. Production planning and coronal stop deletion in spontaneous speech. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1): 15, 1–39.
[29] Tucker, B. V., Ernestus, M. 2016. Why we need to investigate casual speech to truly understand language production, processing and the mental lexicon. *Mental Lexicon* 11(3), 375–400.
[30] Vitevitch, M. S., Luce, P. A. 1998. When words compete: Levels of processing in perception of spoken words. *Psychological Science*. 9, 325–329.
[31] Vitevitch, M. S., Luce, P. A. 2005. Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory and Language* 52, 193–204.
[32] Winter, B. *Applied statistical modeling with R. Generalized linear (mixed) models with linguistic examples*. Submitted.
[33] Wright, R. 2004. Factors of lexical competition in vowel articulation. In: Local, J., Ogden, R., Temple, R. (eds), *Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press, 26–50.
[34] Young, S. J., Kershaw, D., Odell, J., Ollason, D., Valtchev, V., Woodland, P. 2006. *The HTK Book Version 3.4*. Cambridge: CUP.