Phonetic conditioning of word frequency and contextual predictability effects in American English conversational speech

Daejin Kim & Caroline L. Smith

Department of Linguistics, University of New Mexico
daejinkim@unm.edu, caroline@unm.edu

ABSTRACT

Previous studies [1, 3, 5, 6, 8, 11, 19] have shown that words and segments are reduced more when the context consists of more frequent and more predictable words. This study examined these effects as evidenced in the duration and the second formant (F2) of the vowels in monosyllabic CVC words in the Buckeye Corpus [16]. F2 was measured at five timepoints during the entire duration of the vowels. We found that both higher *word* and *bigram* frequency are associated with higher F2 in front vowels (giving a fronter, more extreme quality), while vowel duration shortened or remained the same with higher frequency of both kinds. F2 also increases with higher frequency regardless of the consonantal effect, indicating that higher frequency may make the vowels more resistant to the coarticulatory influence of the consonants. This finding suggests that the more the word is used, the more strongly it is produced in conversation [13, 14, 18].

Keywords: word frequency, contextual predictability, bigram frequency, coarticulation, coarticulatory resistance, corpus phonetics

1. INTRODUCTION

Reduction (or deletion) has been reported in both acoustic and articulatory dimensions due to effects of word frequency [1, 3, 5, 6, 8, 11, 17, 19]. Speakers shorten (or delete) segments more in more predictable words and less in less predictable words. This is because highly predictable words may need less planning and may be produced at a higher speech rate in combination with the preceding and following words, leading to more reduction [12]. Highly frequent words (e.g., *time*) are shorter in duration than their low-frequency homophones (e.g., *thyme*) [2]. In articulation, [11] found that the anterior constriction for [l] was weaker in more frequent words (e.g., *help, milk*) than in their less frequent near-homophones (e.g., *whelp, ilk*). Also, the duration of words and segments is also reduced with increased contextual predictability, also known as *bigram* frequency [17].

On the other hand, the long-term representation of more frequent and predictable words is assumed to include more phonetic details [14, 15]. [21] found that words with high neighborhood density (more frequent with many lexical neighbors) have more expanded vowel space than words with low neighborhood density (less frequent with fewer lexical neighbors). In summary, there is ample evidence that word frequency has the potential to affect the phonetic properties of speech, such as vowel formants. It is less clear in which direction the effect would be, as the research cited above suggests that higher frequency could favor either reduction (centralization) [1] or a more expanded vowel space [21].

Another factor known to affect vowel formants is coarticulation [10]. A variety of phonetic and phonological phenomena are manifested in coarticulation (e.g., [7, 13, 18]); however, previous studies on those effects have not investigated the variation in conversational speech in detail. For example, the value of F2 at the very beginning of the vowel [æ] in monosyllabic English words (e.g. *pad* [pæd]) is lower than it would be in an isolated [æ] due to the coarticulatory effect from the consonant on the vowel. The locus of F2 for [p] is around 600 to 800Hz; therefore, the F2 in the vowel is lowered under the influence of the bilabial consonant [9]. The vowel is likely to show variability in coarticulation from both word and bigram frequency because of their impact on segmental reduction.

This study reports how monosyllabic CVC (C = 7 stop consonants, V = 4 front vowels) words are phonetically realized, considering especially the interaction of coarticulation with word frequency and bigram frequency. Our data source is the American English conversational speech in the Buckeye Corpus [16].

2. PREVIOUS STUDIES

2.1. Segmental reduction effects of word frequency and contextual predictability

A large body of studies found that the more frequent and predictable the word is, the more the speaker reduces the duration of segments [1, 3, 5, 6, 8, 11, 19]. In the spectral dimension, reduction of vowel quality means that the formant values (especially for vowels at the extreme corners of the vowel space, such as [i] and [æ]) become centralized [1], corresponding to a reduced vowel quality. For front vowels, we predict that F2 will decrease (corresponding to a more centralized tongue position) with higher frequency of word or bigram.

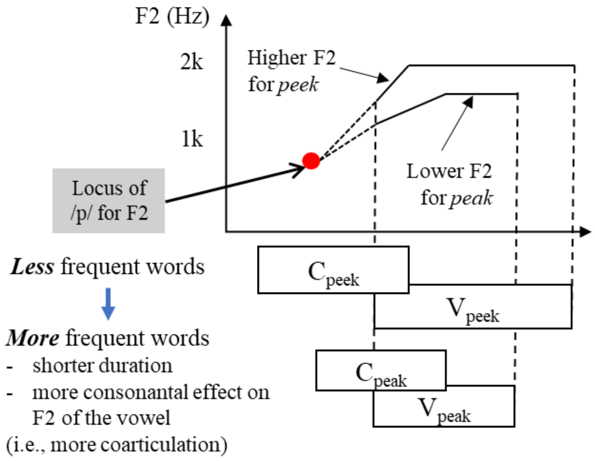
On the other hand, [14, 15] pointed out that some studies found that higher word frequency leads to a stronger representation of the word in the speaker's lexicon. This does not necessarily imply a more extreme articulation of the vowel, just that its identity is stored more robustly.

2.2. The effects of word frequency and contextual predictability on the coarticulation of CV

Both the consonant and the vowel are expected to have a shorter duration in more predictable words than in less predictable ones [1, 5, 6, 8, 19]. In articulation, weaker (smaller amplitude) consonantal and vocalic gestures are expected with more predictable words [11]. The shorter duration of the consonant can be assumed to be a product of the weakening effect resulting from higher predictability. Therefore, the more predictable words may allow more coarticulation between two segments (i.e., coarticulatory aggression) because the segments are more weakly articulated. On the other hand, less predictable words may require a more distinctive acoustic realization, resulting in segments that are less coarticulated with each other (i.e., coarticulatory resistance). However, it is unknown how variation resulting from these frequency-based effects is reflected in conversational speech.

What we anticipate in this study is that each segment in the CV formant transition of less frequent and predictable monosyllables will be more clearly realized than in more frequent and predictable words. We predict less gestural overlap, resulting in formants closer to the values they would have in an isolated vowel. Following this line of reasoning, a vowel with lower frequency may have F2 at onset different from its value at the midpoint. For example, *peak* [pik] is more frequent than its homophone *peek*. As shown in Figure 1, the F2 of [i] in *peak* may be lower than that in *peek*, indicating that the degree of coarticulation is affected by word frequency. We expect less frequent words (e.g., *peek*) to need more distinctive articulation with lengthening of duration, while the more frequent homophone *peak* will have segments that exhibit reduction in quality with shortening of duration.

Figure 1: The prediction of the degree of coarticulation of high frequency *peak* vs. low frequency *peek* in the F2.



This may result in the [p] in *peak* having a larger effect on the vowel formants than in *peek*.

3. METHODS

The data for this study is taken from the phonetically transcribed conversational dialogues in the Buckeye Corpus [16]. Using a phonetically annotated conversational speech corpus enables us to examine the phonetic correlates of coarticulation with a huge variety of contextual variability. The 20 younger (10 male and 10 female) speakers in the corpus were analyzed. Each conversation has a duration from 30-60 minutes. For this study, tokens of 71 different monosyllabic <u>CV</u>C words with combinations of four consonant types at the onset (<u>C</u> = Bilabial [p, b], Alveolar [t, d], Velar [k, g], and Glottal [h]) and the four front vowels (V = iy [i], ih [ɪ], eh [ɛ], ae [æ]) were extracted from these dialogues.

To observe the effects of word and bigram frequency on the phonetic properties of these CVC monosyllabic words, the second formant (F2) and the vowel duration were analyzed. F2 values were extracted at five timepoints (1 – 5) during the vowels, from 10ms after the beginning of the vowel (1), at 25% of the duration of the vowel (2), at the vowel midpoint (50%) (3), at 75% of vowel duration (4), and 10ms before the end of the vowel (5). The timepoint measure was statistically analyzed as a continuous variable. This timeline measure may enable us to analyze the degree of coarticulatory influence of the consonant on the vowel more precisely than if measures were made only at the vowel midpoint. Tokens with second formant bandwidth above 700Hz were excluded, and tokens with vowel duration less than 50ms were excluded. In total, 5988 tokens were analyzed. Both the F2 and the duration of the vowels were normalized (*z*-scored) by each speaker's average[i].

For statistical analysis, we performed a linear mixed effects analysis [2] to observe the effects of word (WF) and bigram (BF) frequency on the F2 and the vowel duration. We included a random intercept for subjects in the models predicting WF and BF effects on F2 and vowel durations. Note that only between-speaker differences in F2 and vowel duration were statistically significant; therefore, speaker was the only random effect in the models. The independent variables in the analysis were:

- **Word frequency**: Raw word frequency (WF) is taken from SUBTLEX$_{US}$ Corpus [4]. It indicates the occurrence of the word per million words and is converted into the logarithmic scale.
- **Preceding Bigram frequency**: Preceding bigram frequency (BF) measures the frequency of occurrence of the string consisting of the target word and the word preceding it. This was calculated as a ratio over the entire Buckeye Corpus by dividing the number of occurrences of the word pairs by the total number of occurrences of the target word, then converting to a logarithmic scale. Since word and bigram frequency measures are not correlated (r^2 = -0.08), we included the BF and WF effects together in the model [19].
- **F0 difference from the previous word:** Prosodic factors are critical in predicting the degree of coarticulation and its acoustic correlates. For example, a vowel with pitch accent or stress has more coarticulatory resistance to adjacent segments ([7]). To account for variation due to prosodic factors, for each word we calculated the f0 difference between the target vowel and the vowels in the preceding word. However, this had little effect on the F2 and vowel duration in the statistical model.

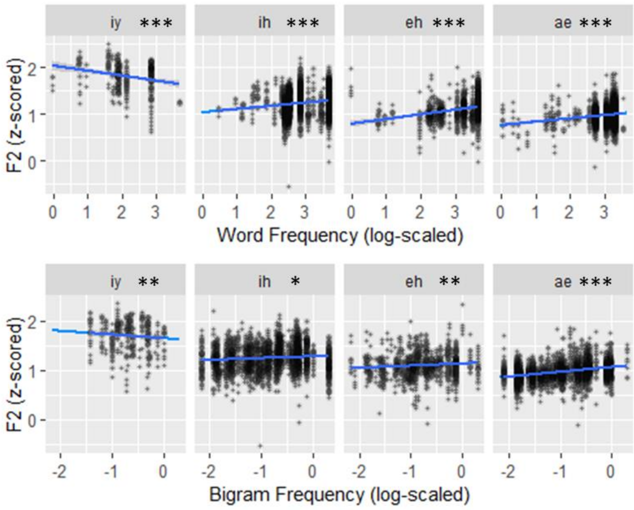
4. RESULTS

4.1. F2 (vowel front-back)

The model[ii] estimates the effects of WF and BF on F2 with the different front vowels and consonants. Figure 2 plots the increase in F2 values with higher WF and BF that was observed except in the 'iy' [i] vowel, where F2 decreases with higher word frequency. (WF on 'iy': *β* (estimate) = -0.21, SE (standard error) = 0.07, *p* < .001; WF on 'ih' [ɪ]: *β* = 0.18, SE = 0.03, *p* < .001; WF on 'eh' [ɛ]: *β* = 0.21, SE = 0.03, *p* < .001: WF on 'ae' [æ]: *β* = 0.18, SE = 0.03, *p* < .001).

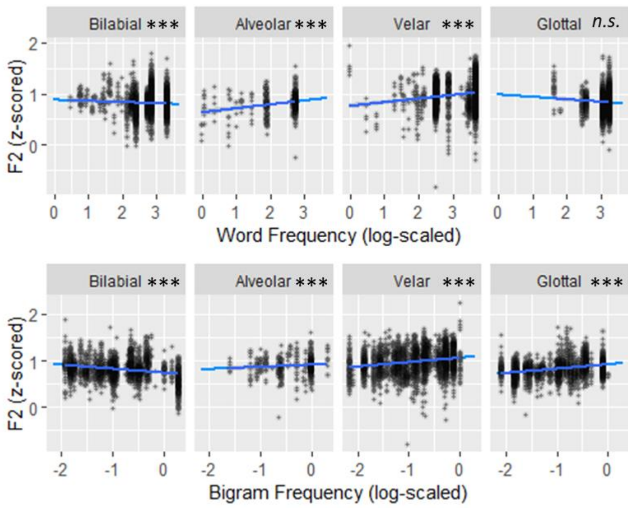
Bigram frequency (BF) also increases the F2 vales of the vowel except for the vowel 'iy'. (BF on 'iy': *β* = -0.24, SE = 0.05, *p* > 0.1; BF on 'ih': *β* = 0.09, SE = 0.05, *p* < .001; BF on 'eh': *β* = 0.09, SE = 0.05, *p* < .001: BF on 'ae': *β* = 0.18, SE = -0.02, *p* > 0.01). This indicates that except for the vowel 'iy', speakers enhance the phonetic characteristics of the vowel (i.e., [+front]) with a more forward tongue position, in words with higher frequency and contextual predictability. This finding is in the same direction as reported by [18] but opposite to the findings from some previous studies, such as [1, 3, 5, 6, 8].

Figure 2: Scatterplots of F2 for the WF and BF effects by vowel type. Each regression line fits the model indicating the directionality of those effects. Statistical significance of the effects is indicated as *** (*p* <.001), ** (*p* <.01), and * (*p* < .05) for each vowel.



Our results also showed WF and BF effects that are different from previous studies and our expectations (Figure 3). For all front vowels with alveolar and velar consonants, the F2 increases (or remains the same) with both WF and BF effects (WF with Alveolar: *β* = 0.09, SE = 0.02, *p* < .001; WF with Velar: *β* = 0.09, SE = 0.02, *p* < .001; BF with

Figure 3: Scatterplots of F2 for the WF and BF effects by consonant type. Each regression line fits the model indicating the directionality of those effects. Statistical significance of the effects is indicated as *** (*p* <.001) and *n.s.* (*p* > .1) for each consonant type.

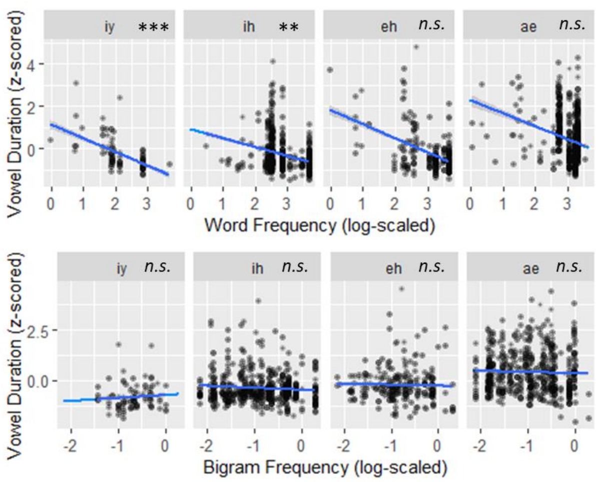


Alveolar: β = 0.09, SE = 0.02, p < .001; BF with Velar: β = 0.09, SE = 0.02, p < .001; BF with Glottal: β = -0.02, SE = 0.05, p > 0.1). In contrast, F2 did not increase in vowels preceded by bilabial consonants (WF with Bilabial: β = -0.21, SE = 0.07, p < .001; BF with Bilabial: β = -0.24, SE = 0.05, p < .001). Note that these effects point to the difference due to WF and BF effects; F2 is naturally lower in the context of bilabials than alveolars, but that effect is not reported here. We predicted that F2 would decrease more with higher WF and BF. Instead, we observed that F2 *increases* (or maintains) with higher WF and BF in most contexts. This indicates that some of these vowels show less evidence of coarticulation, being resistant to coarticulatory influence from the consonant at the onset, though there are some exceptions.

4.2. Vowel duration

Unlike the WF and BF effects on F2, the statistical model predicting vowel duration shows that higher WF shortens all vowels, and that higher BF is associated with non-significant shortening except for the words with 'eh' [ɛ] vowel. As shown in Figure 4, F2 decreases more with higher WF (WF on 'iy' [i]: β = -0.65, SE = 0.19, p < .001; WF on 'ih' [ɪ]: β = -0.40, SE = 0.08, p < .001; WF on 'eh' [ɛ]: β = 0.02, SE = 0.08, p > 0.1: WF on 'ae' [æ]: β = 0.08, SE = 0.03, p > .1). Higher BF has little effect on vowel duration (BF on 'iy': β = 0.10, SE = 0.19, p > 0.1; BF on 'ih': β = -0.25, SE = 0.08, p = 0.06; BF on 'eh': β = 0.13, SE = 0.08, p > 0.1; BF on 'ae': β = 0.53, SE = 0.03, p = 0.08).

Figure 4: Scatterplots of vowel duration for the WF and BF effects by vowel type. The regression lines fit the model indicating the directionality of those effects. Statistical significance of the effects on the vowel duration is indicated as *** (p <.001), ** (p <.01), and *n.s.* (p > .1) for each vowel.



5. DISCUSSION & CONCLUSION

Previous studies of lexical and contextual effects in speech production have found that speakers decrease the duration of words and segments with higher predictability [1, 3, 5, 6, 8, 19]. Our measure of BF gets at the issue of predictability: a higher bigram frequency means the target word frequently follows the same preceding word, creating a more predictable string. However, what our study found is that speakers strengthen the vowel quality by giving it a more extreme quality with higher WF and BF. In an articulatory dimension, it is possible to assume that they strengthened the tongue movement by moving further forward. As [14, 15] assumed, the more the word is experienced in production, the more specific phonetic detail is stored in the lexicon. Shortened duration, on the other hand, may be a consequence of faster access to the lexicon with higher predictability.

In terms of coarticulation, F2 increases despite the expected effects of the adjacent consonants that are likely to influence F2. This finding suggests that with higher frequency of the word and the context, the vowel is more resistant to coarticulation, consistent with findings in [18]. Although speakers shorten (or maintain) the duration of the vowels, during the part of the vowel close to the source of coarticulation (the consonant), the F2 for these front vowels increases with increased frequency (as indicated in the bottom part of Figure 1). This tells us that the vowel shows resistance in coarticulation and strengthening in quality, suggesting that frequency effects at the lexical and contextual level are not entirely reduction.

Speech production in conversation may have a huge amount of variation in phonetic and phonological variations compared with experimental settings. However, even with the reduction of the vowel and coarticulation from the onset consonants, the F2 increases with higher frequency and predictability in context regardless of the consonantal effect to the vowel. This result suggests that these fine-grained phonetic effects may be controlled by the speaker [20].

In conclusion, the effects of word frequency and contextual predictability are associated with a stronger articulation by enhancing the phonetic properties of segments regardless of the reduction of segmental duration. In terms of coarticulation, the vowel is more resistant to coarticulation with higher frequency, supporting this paper's finding of phonetic conditioning pattern of these frequency-based effects.

6. REFERENCES

[1] Aylett, M., Turk, A. 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *J. Acoust. Soc. Am.*, *119*(5), 3048-3058.

[2] Bates, D., Mächler, M., Bolker, B., Walker, S. 2014. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

[3] Bell, A., Brenier, J. M., Gregory, M., Girand, C., Jurafsky, D. 2009. Predictability effects on durations of content and function words in conversational English. *J. Memory and Lang.*, *60*(1), 92-111.

[4] Brysbaert, M., New, B. 2009. Moving beyond Kucera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English. *Behavior Research Methods, 41*(4), 977-990.

[5] Bush, N. 2001. Palatalization in English. In Bybee, J., P. Hopper (eds), *Frequency and the Emergence of Linguistic Structure*, 255-280.

[6] Bybee, J. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Lang var and change*, *14*(3), 261-290.

[7] Cho, T., Kim, D., Kim, S. 2017. Prosodically-conditioned fine-tuning of coarticulatory vowel nasalization in English. *J. Phon.*, *64*, 71-89.

[8] Gahl, S. 2008. Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474-496.

[9] Kent, R. D., Read, C. 1992. *The acoustic analysis of speech*. San Diego: Singular Publishing Group.

[10] Kühnert, B., Nolan, F. 1999. The origin of coarticulation. In W. Hardcastle, N. Hewlett (eds), *Coarticulation: Theory, data and techniques*, 7-30.

[11] Lin, S., Beddor, P., Coetzee, A., 2014. Gestural reduction, lexical frequency, and sound change: A study of post-vocalic/l. *LabPhon*, *5*(1), 9-36.

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

[13] Perillo, S., Bang, H. Y., Clayards, M. 2015. Locus equation metrics as an index of coarticulation resistance: The effect of prosodic prominence. In *Proc. Meet on Acoust. 25*(1), 60007.

[14] Pierrehumbert, J. 2002. Word-specific phonetics. *LabPhon*, *7*, 101-139.

[15] Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, J. P. J. Hopper (eds), *Frequency and the Emergence of Linguistic Structure.*, *45*, 137-158.

[16] Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., Raymond, W. 2005. The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability. *Sp Comm.*, *45*(1), 89-95.

[17] Raymond, W. D., Brown, E. L. 2012. Are effects of word frequency effects of context of use? An analysis of initial fricative reduction in Spanish. *Frequency effects in language learning and processing*, *1*, 35-52.

[18] Scarborough, R., 2004. *Coarticulation and the structure of the lexicon* (PhD dissertation. UCLA).

[19] Schuppler, B., van Dommelen, W. A., Koreman, J., Ernestus, M. 2012. How linguistic and probabilistic properties of a word affect the realization of its final/t: Studies at the phonemic and sub-phonemic level. *J. Phon.*, *40*(4), 595-607.

[20] Solé, M.-J. Controlled and Mechanical Properties in Speech. In Solé, M.-J., Beddor, P., Ohala, J. (eds), *Experimental Approaches to Phonology*. 302-321. Oxford Univ. Press.

[21] Wright, R. 1997. Lexical competition and reduction in speech: A preliminary report. *Res on spoken lang processing prog report*, *2*, 471-485.

---

[i] We normalized (*z*-scored) the data with all types of vowels, including front, mid, back vowels, and with both F1 and F2 values. This paper presents the results from the F2 values of front vowels only.

[ii] All mixed effects models reported here were analyzed with the normalized (*z*-scored, logarithmic-scaled) values to ensure the linearities in the prediction. Models are also all statistically significant from the models with less variables and with more variables.