A Reanalysis of the Voicing Effect in English: With implications for featural specification

Rebecca L. Morley and Bridget J. Smith

The Ohio State University 1712 Neil Ave.

Columbus OH 43210

614 292-4052

morley.33@osu.edu

Abstract

The voicing effect is among the most studied and most robust of phonetic phenomena. Yet there remains a lack of consensus on why vowels preceding voiced obstruents should be longer than vowels preceding voiceless obstruents. In this paper we provide an extensive review of roughly seventy years of literature, an analysis of the voicing effect in a corpus of natural speech, and production data from a metronome-timed word repetition study. From this evidence we conclude that: vowel duration differences follow from consonant duration differences; the voicing effect is largely limited to words of especially long duration; and preceding vowel duration does *not* reliably cue obstruent voicing under the following circumstances: when obstruent voicing or duration cues conflict; for lax or unstressed vowels; and for most conversational speech. We show that this behavior can be modeled using a competing-constraints framework, where all segments resist expanding or compressing past a preferred duration. Inherent segment elasticity determines the degree of resistance, and duration is ultimately determined by the interaction of these segmental constraints with constraints on target rhyme duration (as a measure of speaking rate), and a preferred C/V duration ratio. Because these constraints are implemented as continuous probability distributions, under- and over- shoot of target duration is possible, and explicit temporal compensation is not required. This account of the voicing effect has a number of implications for phonological theory, especially the central role that the concept of prominence plays in the analysis of underlying features.

Keywords: voicing effect, vowel lengthening, final lengthening, temporal compensation, enhancement, Articulatory Phonology

A Reanalysis of the Voicing Effect in English: With implications for featural specification

The terms “vowel lengthening” and “voicing effect” are used to refer to the empirical finding that vowels preceding voiced obstruents are longer than those preceding voiceless obstruents (e.g., Sweet, 1880; House and Fairbanks, 1953; Denes, 1955; Peterson and Lehiste, 1960; House, 1961; Sharf, 1962; Chen, 1970; Raphael, 1972; Klatt, 1973; Lisker, 1974; Raphael, 1975; Raphael et al., 1975; Umeda, 1975; Klatt, 1976; Port, 1976; Fox and Terbeek, 1977; Javkin, 1977; Lisker, 1978; Derr and Massaro, 1980; Fitch, 1981; Walsh and Parker, 1981; Crystal and House, 1982; Krause, 1982; Port and Dalby, 1982; Ohala, 1983; Hillenbrand et al., 1984; Luce and Charles-Luce, 1985; Lisker, 1986; Van Summers, 1987; Kluender et al., 1988; Fischer and Ohde, 1990; De Jong, 1991; Laeufer, 1992; Crowther and Mann, 1992; Braunschweiler, 1997; Smith, 2002; De Jong, 2004; Kulikov, 2012; Ko, 2018; Tanner et al., 2019; Sanker, 2019; Coretta, 2019; Beguš, 2017). Voicing effects have been documented in a number of different languages. However, it is generally agreed that English (in many of its varieties) exhibits one of the strongest such effects, with pre-voiced vowels up to 50% longer than their pre-voiceless counterparts (e.g., Chen, 1970; Harris and Umeda, 1974; Mack, 1982). English speaking listeners also exhibit a robust categorical perception effect for final voicing based on preceding vowel duration alone (e.g., Raphael, 1972; Crowther and Mann, 1992; Klatt, 1976; Hillenbrand et al., 1984; Denes, 1955). Preceding vowel duration has been described as the most reliable cue to voicing on final obstruents (Raphael, 1972; Raphael et al., 1975; Luce and Charles-Luce, 1985). Because stops are often unreleased in word-final position, it has also been suggested that a sound change has occurred (or is underway) in which the contrastive relationship between words like “bad” (bæd) and “bat” (bæt) has shifted away from the final obstruent itself, to be expressed in the duration of the preceding vowel (bæˑt̚ vs. bæt̚), at least in phrase-final position (e.g., Klatt, 1976).

In this paper, however, we will argue that the primary cue to obstruent voicing in coda position is the duration of the obstruent itself. Short obstruents are perceived as members of the phonologically voiced category, and long obstruents, as members of the voiceless. Because the long/short distinction is relative, preceding vowel duration, as an indicator of speaking rate, affects the perception of voicing. We find that the duration of the preceding vowel is better predicted by the following obstruent duration than by its voicing (whether phonetic or phonological). We argue that the two phonological categories of stop are distinguished by their inherent elasticity: the less elastic a segment, the more it resists both lengthening and shortening.

The paper is organized as follows. In the next section we provide a literature review on the voicing effect and related phenomena. Section 2 contains a corpus study on American English. In Section 3, predictions of our hypothesis are tested using a variable-rate production task. In Section 4 we model the results using continuously violable constraints on duration at the segment and syllable level. Our hypothesis is elaborated in Section 5, where we account for the perceptual side of the voicing effect. We summarize and conclude in Section 6, where we discuss the implications of the present work for theories of phonological contrast.

1 Background

Despite the large amount of research on the phenomenon, the underlying source of the voicing effect is unknown, with little consensus on what acoustic or articulatory properties give rise to the observed duration differences. Belasco (1958), and Delattre (1962), both argue that there is a trade-off of effort between the vowel and consonant: strong (voiceless) consonants are accompanied by weak (shorter) vowels. However, Moreton (2004) and Schwartz (2010) argue essentially the opposite: that it is the spread of "fortisness" that shortens the preceding vowel. It has also been claimed that careful, and therefore slower (or simply earlier, see Klatt 1976), movements of the vocal cords are required to avoid spontaneous voicing under reduced pressure (Halle and Stevens, 1967). However, clear evidence of differences in energy, effort, or precision between voiced and voiceless obstruents has not been forthcoming.

On the auditory side, Kluender et al. (1988), and also Jessen (2001), suggest that vowel lengthening is an enhancement effect, reinforcing the length differences of the obstruents, and thus the voicing contrast. Javkin (1977) posits that vowels are consistently perceived as longer before voiced obstruents because listeners mis-attribute glottal pulsing to the end of the vowel.

However, there seems to be little evidence to support the latter hypothesis, and enhancement explanations are unable to account for why it is preceding vowel length, and not obstruent length, degree of voicing, presence of audible release, or aspiration that are used to enhance the contrastiveness of the obstruents themselves. More recently, Sanker (2020) has proposed an explanation based on the interaction between acoustics and articulation: a subset of the features in the vowel that are affected by the voicing of the following obstruent (spectral tilt, and intensity contour) also affect perception of vowel duration, presumably for unrelated articulatory reasons. Thus, in the presence of those cues, independent of the presence of a following obstruent, listeners perceive the vowel as being longer/shorter than some baseline duration. This proposal hinges on the source of the duration percept being independent of the voicing, which has not been established, since no articulatory explanation for why spectral tilt and intensity contour affect perceived duration has been proposed.

1.1 Production

The above hypotheses encounter more problems when the voicing effect is examined more closely. The articulation-based hypotheses that rely on actual vocal fold vibration cannot account for the fact that voicing effects occur even when voiced obstruents are phonetically devoiced (e.g., Walsh and Parker, 1981; Chen, 1970; Fox and Terbeek, 1977). A universal basis for the effect is also called into question by the apparent absence of a lengthening effect in certain languages (Flege, 1979; Hillenbrand et al., 1984; Keating, 1979, 1985).[1] Even in English, with one of the most robust voicing effects measured, durational differences are not found in all contexts. Production studies typically consist of either word lists or brief sentences read by participants in a laboratory setting. In sentence contexts, the target words are often in absolute final position. Such words are also typically monosyllabic, which entails that the target vowel receives primary stress. When some or all of these factors are varied, the voicing effect can be significantly reduced, or disappear altogether: in phrase-medial position (versus phrase-final)

[1] Although it should be kept in mind that the cross-language comparisons are not always of like items in these studies.

(Umeda, 1975; Smith, 2002; Crystal and House, 1988; Luce and Charles-Luce, 1985), polysyllabic words (versus monosyllabic) (Umeda, 1975; Port, 1981; Klatt, 1973), lax vowels (versus tense) (Crystal and House, 1988; Luce and Charles-Luce, 1985; Peterson and Lehiste, 1960), unstressed vowels (versus stressed vowels) (Van Summers, 1987; De Jong, 2004), and fast speaking rates (versus slow speaking rates) (Port, 1976; Smith, 2002; Ko, 2018).

1.2 Compensation

Obstruent duration differences mirror vowel duration differences: they are small or non-existent in the contexts in which vowel duration differences are small or non-existent (Luce and Charles-Luce, 1985; Crystal and House, 1982; Miller et al., 1986),[2] and they are large where large vowel duration differences are found, and in the opposite direction (e.g., Klatt, 1976; Umeda, 1975; Miller and Volaitis, 1989; Chen, 1970; Luce and Charles-Luce, 1985). The inverse correlation between differences in obstruent duration and differences in preceding vowel duration was noted early on (Kozhevnikov and Chistovich, 1965; Catford, 1977). However, temporal compensation as an explanation for the voicing effect has been explicitly considered and rejected on a number of separate occasions (Keating 1985; Chen 1970; Braunschweiler 1997). These rejections are largely based on the assumption that temporal compensation should be total, or near total,[3] meaning that syllables would all be the same length.

Although syllable-level isochrony was originally hypothesized to apply in so-called "syllable-timed" languages like English (e.g., Pike, 1945), and to be the source of a number of apparently compensatory effects, it has become clear that uniform timing for syllables is not consistently enforced in English[4] or in any other language that has been investigated (see

[2] Calculated either in absolute terms, or as percentages

[3] There are, in fact, a handful of studies that report vowel duration differences that are very close to closure duration differences across minimal pairs (in English: Lisker, 1957; Sharf, 1962; Davis and Van Summers, 1989; in Polish: Coretta (2019); in Georgian:Beguš (2017)). However, across studies, the measured stops were in word-medial position, or in polysyllabic words. Most were produced phrase-medially. Some stops appeared in post-stress position; and for some stimuli, there may have been a syllable boundary between the consonant and the vowel. For all these reasons, the effect sizes were quite small, with duration differences for both vowels and stop closures ranging between 8 and 35 ms.

[4] Syllables with low vowels are generally longer than those with high vowels (Peterson and Lehiste, 1960); syllables

Krivokapić (2020) for a review).

1.3 Competition

More recent work on temporal compensation is situated within Articulatory Phonology (AP), which does not assume isochrony at any level. This is appealing for "compensatory" phenomena that range widely in their degree (e.g., Elert, 1965; Kristoffersen, 2000; Kavitskaya, 2002; Munhall et al., 1992; Kim and Cole, 2005). In AP, articulatory units of various sizes are modeled as harmonic oscillators with different characteristic frequencies (e.g., Browman and Goldstein 1990; Nam and Saltzman 2003; Saltzman et al. 2008; Browman and Goldstein 1988; O'Dell and Nieminen 1999). Phasing relationships between such articulatory units are derived from oscillator coupling, resulting in a system with its own characteristic frequency. The same result can be derived from a competing constraints model in which none of the individual constraints on preferred frequencies can be perfectly satisfied, and a "compromise" frequency is adopted. Our model of the voicing effect is based on this approach.[5]

1.4 Intrinsic Duration

In our proposed model it is preferred durations at the segment level that drive the voicing effect. We posit that something similar is at work in so-called prominence-based compensation, which occurs between two syllables of inherently different durations within the same word. Final lengthening associated with phrasal boundaries is typically strongest for the segment closest to the boundary, and extends only as far as the onset of the final syllable in most cases (Turk and Shattuck-Hufnagel, 2007; Cambier-Langeveld, 1997; Berkovits, 1993; Hofhuis et al., 1995; Campbell, 1992; Port and Cummins, 1992). However, Cambier-Langeveld (1997; 2000) show that, in Dutch, the penultimate syllable of the final word also sometimes experiences significant

with tense vowels tend to be longer than syllables with lax vowels (Peterson and Lehiste, 1960; Sharf, 1962); stressed syllables are longer than unstressed syllables (De Jong, 2004).

[5] Browman and Goldstein (1986) themselves adopt one of the phonetic explanations for the voicing effect: because voiceless stops require extra glottal opening and closing, their preceding vowels are shorter.

lengthening. This happens only when the final syllable is unstressed, or contains a schwa vowel (see also, Turk and Shattuck-Hufnagel, 2007,Katsika (2016) for similar results).

In these studies, the characteristically shorter duration of unstressed vowels seems to prevent them from lengthening to a degree sufficient to satisfy the requirements of phrase-final lengthening. The voicing effect can be described in similar terms: lengthening (also often due to a phrase-final boundary) “shifts” to earlier segments (the vowel) when the final segment (the voiced obstruent) cannot be lengthened sufficiently.[6]

1.5 Elasticity

Our model assigns a characteristic elasticity to each segment, which mediates the degree to which the segment resists pressures to lengthen or shorten from its preferred duration (see also Cambier-Langeveld (2000) and Miller (1981)). The concept of elasticity is related to the concept of spring stiffness in Articulatory Phonology (e.g. Browman and Goldstein 1986): a force of the same size will cause a greater perturbation to a spring with a smaller stiffness parameter, resulting in a longer duration. However, elasticity differs from spring stiffness in important ways: it applies to phonemes, which are not part of the inventory of timing units within AP, and not just to phonemes as a class, but to individual phonemes.

What is crucial in our model is that elasticity and duration determine the proportion of the syllable that each segment comprises (see Campbell 1992). Our Expandability Hypothesis is defined in (1).

(1) The Expandability Hypothesis

All segments have a characteristic elasticity that determines their resistance to lengthening

Resistance to lengthening increases with increasing duration for all segments

Lower elasticity equates with a more rapid increase in resistance

[6] Although Munhall et al. (1992) suggest that differences in vowel duration preceding voiced versus voiceless obstruents can be explained by differences in the phasing of the two consonants with respect to the preceding vowel, they do not actually provide any evidence in support of this view.

Relative resistance determines the distribution of duration across the syllable

Modeling voiceless obstruents as high elasticity, and voiced obstruents, as low elasticity, we will show that the Expandibility Hypothesis parsimoniously accounts for the production data on the voicing effect. Crucially, the Expandability Hypothesis is also consistent with the perception data.

1.6 Perception

It turns out that vowel duration is not the only cue, and may not even be the main cue, to the voicing distinction. Wardrip-Fruin (1982) demonstrates that when preceding vowel duration conflicts with either formant transition cues, or actual vocal fold vibration, the latter dominates. Hogan and Rozsypal (1980) also report that, for certain voiceless-final words, lengthening the vowel does not change the percept to voiced, but produces no effect, or results in stimuli that sound unnatural. Revoile et al. (1982), using naturally produced stimuli, find that the identification of voiced stops is most strongly disrupted by removing vowel offset cues (see also O'Kane, 1978; Nittrouer, 2004), while the identification of voiceless stops is most strongly disrupted by removing the release burst. Similarly, Repp and Williams (1985) find that the addition of a release burst to otherwise ambiguous stimuli reduces voiced responses. Changes to vowel duration, on the other hand, have little effect on voicing perception in their study.

It is our hypothesis that the perception results are based primarily on obstruent duration. It was established quite early on that the perceptual boundary between the fricatives /s/ and /z/ in final position is dependent on both consonant and vowel duration (Denes, 1955). See also, Raphael (1981) and Repp and Williams (1985). In fact, the literature on voicing in word-medial position standardly describes the perceptual boundary in terms of the ratio of closure duration to preceding vowel duration (e.g., Port and Dalby, 1982; Port, 1979, 1981; Lisker, 1957). The C/V ratio effectively normalizes stop duration relative to estimated speaking rate. This is exactly what we believe occurs in final position, with final lengthening accounting for the increased size of the effect.

2 A Corpus Study

In this section we provide an in-depth analysis of the voicing effect in conversational speech, using data from the Buckeye Corpus (Pitt et al., 1997). Although the corpus is not balanced, it provides much more data, and a larger range of speaking rates and contexts than any single laboratory experiment. A corpus study allows us, first, to quantify the voicing effect in actual usage. Secondly, it allows us to probe more deeply into the factors that affect the realization of the effect. Conversational styles of speech are expected to exhibit considerable reduction in the realization of individual words, some component sounds of which may be entirely missing (e.g., Harris and Umeda, 1974; Johnson, 2004; Jurafsky et al., 1998). This reduction could neutralize small differences in duration that result from an underlying voicing effect. And previous studies with read scripts have shown a reduced voicing effect in comparison to single sentence or word list productions (Crystal and House, 1982, 1988). We find that there is an inconsistent effect of voicing that is dependent on model structure. When consonant duration is added to the model the voicing effect either goes away, or switches direction. Consonant duration, on the other hand, is consistently negatively correlated with vowel duration. Furthermore, both effects are found to participate in interactions with speaking rate and frequency, exhibiting a dependence on absolute duration.

2.1 The data

The Buckeye Corpus consists of segmented and transcribed sound files. These are taken from interviews, each lasting about an hour, with 40 different speakers, all middle-class and Caucasian, who are also natives of central Ohio. Intertranscriber reliability of the phonetic symbols for stops and fricatives was reported for a sample of the Buckeye Corpus at 91.2% and 92.9%, respectively. For the unanimously transcribed subset of this sample, segmentation boundaries differed an average of 16 ms. (Pitt et al., 2005). However, Raymond et al. (2002) report a difference in segmentation agreement for shorter versus longer phones. 73% of phones that were longer than average agreed within 20% of the average length of the two phones on either side of the segment boundary, whereas only 50% of phones that were shorter than average agreed within 20%. Shorter phones were thus proportionally less consistently transcribed than longer phones. In the absence of a consistent bias in the placement of the boundary, such errors could wash out a small voicing effect. Given that the voicing effect is expected to be larger for longer durations, however, this is unlikely to affect the outcome significantly. The segmentation of the vowel and final consonant are inherently negatively correlated; an error in which the vowel duration is longer will also produce an error in which the final obstruent is shorter. However, such ambiguity is more likely to arise with voiced than with voiceless stops. Thus, we would expect such errors to inflate any voicing effect. This is not what we find however.

From the Buckeye Corpus we extracted all monosyllabic words of the form (C)onsonant-(V)owel-(C)onsonant ending with one of the following obstruents: voiced (d,b,g,z,ʒ,v) or voiceless (t,p,k,s,ʃ,f). CVC words were selected because they were expected to show the largest voicing effect. Complex onsets were excluded to eliminate potential variability. No nasalized or rhotacized vowels were included, to be sure that each word had exactly three underlying segments. Only tokens that were both phonemically and phonetically CVCs were included. For example, tokens of "past" realized as [pæs], and tokens of "allowed" realized as [la͡ʊd] were both excluded. Because the transcription of the corpus is quasi-phonetic, we constructed a dictionary of citation forms to ensure that the phonological voicing category was correctly assigned to each word. Because there were no words ending in voiced dental fricatives, those ending in voiceless dental fricatives were also removed. The vowel /o͡ɪ/ was also excluded for reasons of data sparsity. 20.3% of the stops in the remaining data were transcribed as glottalized (tq), which could represent a glottal stop or unreleased stop with glottalization on the vowel, but less than 1% of those were underlyingly voiced, so all such tokens were removed from analysis. Affricates were excluded due to the possibility that they might straddle a word boundary.

In Figure 1 raw vowel durations for the set of CVC word tokens used in the following analyses are plotted as a function of the voicing feature of the final obstruent. The density plot on the right suggests that there is a very small effect of voicing at the longest durations. However, the actual counts given in the left panel show that there are never more voiced than voiceless tokens at any duration. This is due to the fact that there are considerably more word tokens with (phonemically) voiceless coda obstruents (over twice as many as voiced tokens, although there are more voiced than voiceless fricative tokens. See Appendix A). Vowels preceding voiceless obstruents have a slightly longer mode than those preceding voiced obstruents, and at the longer durations (above 175 ms.), the *relative* proportion of the pre-voiced distribution is larger than the pre-voiceless. For the most part, however, the two distributions are completely overlapped, showing no transparent voicing effect.

**** FIGURE 1 ABOUT HERE *******

2.2 Model Factors

If a voicing effect does exist in these data, it is masked by factors that affect vowel duration more strongly. The following factors, each of which is known to affect segment duration, are included in the statistical model of vowel duration. Because the analysis was limited to CVC words, stress and word length are not included.

- INHERENT VOWEL CLASS: Tense and lax vowels in English are differentiated in part by duration. /ɪ, ɛ, ʊ, ʌ/, all lax vowels, are reliably shorter than their tense counterparts (e.g., Peterson and Lehiste, 1960; Klatt, 1976; Stevens and House, 1963). /æ/, although technically lax, has much longer durations than any other lax vowel (e.g., Hillenbrand et al., 2000; Crystal and House, 1988), and is grouped with other inherently long (tense) vowels (see also Fulop and Scott 2021; Rositzke 1939. Reduced or absent voicing effects have been reported for both unstressed and lax vowels (Umeda, 1975; Crystal and House, 1982; De Jong, 2004). To capture this length difference, Vowel Class is modeled as a factor with 2 levels: Short (ɪ, ɛ, ʊ, ʌ), and Long (all other vowels, namely, i,e,u,ɑ,o,ɔ,æ,ɑɪ,ao), coded as 1, and -1, respectively.

- VOWEL HEIGHT: Because high vowels tend to be shorter than low vowels, this can affect the realization of the voicing effect. Vowel height is a factor with 2 levels : high (i,u,ɪ,ʊ), and non-high (all other vowels), coded as -1, and 1, respectively.

- SPEAKING RATE DEVIATION: The z-scored average difference between expected and observed duration was used as a proxy for rate difference (see Gahl et al., 2012; Priva, 2017). Mean segment durations by speaker were taken as the expected value for each phoneme. This was computed over all tokens (regardless of word position or phonological context) and for all words in the Buckeye Corpus (not just the CVC words used in the analysis). The difference was calculated for each segment within a word, and then averaged. Because this is actually a duration measure, a positive difference indicates that the individual segments within the word are generally longer than their average durations, and thus that the speaking rate is slower than average. Speaking rate deviation is modeled as a continuous variable.

- WORD FREQUENCY: More frequently used words generally have shorter durations than less frequently used words, and both vowels and consonants within those words are affected (e.g., Jurafsky et al., 2001; Fidelholtz, 1975; Fosler-Lussier and Morgan, 1999; Hooper, 1976; Pluymaekers et al., 2005). Function words, generally the most frequent and the most contextually predictable words, are consistently shorter than content words (Bell et al., 2009; Umeda, 1975). Because the difference in frequency between content and function words is several orders of magnitude, Zipf scores, $log_{10}(Frequency)$, were used. Word frequencies were supplied as counts per million from the SUBTLEX corpus (Van Heuven et al., 2014). Log-frequency is modeled as a continuous variable.

- PHRASAL POSITION: Prosodic boundaries have the effect of lengthening adjacent segments. The greater the number of nested phrases marked by the boundary, the greater the degree of lengthening, and the further its spread (Oller, 1973; Wightman et al., 1992; Fougeron and Keating, 1997; Byrd and Saltzman, 2003). Because the Buckeye Corpus is not annotated for syntactic boundaries, tokens were classified only as pre-pausal or non-pre-pausal, based on the end of a transcribed utterance. Pre-pausal position is expected to show the largest lengthening effects (see, e.g., Crystal and House, 1988; Klatt, 1975). The following tags were used to identify a phrasal boundary: SIL (silence), E_TRANS (end of phonetic transcription), IVER (interviewer speaking), VOCNOISE (non-speech sound such as a cough, or laugh). Position is modeled as a factor with 2 levels: phrase-final and non-phrase-final, coded as 1 and -1, respectively.

- PHONETIC VOICING: Phonetically voiced segments exhibit acoustic evidence of voicing, as transcribed by corpus annotators. Phonetic voicing is modeled as a 2 level factor: voiced, and voiceless, coded as 1 and -1, respectively.

- PHONEMIC VOICING: Phonemic voicing refers to the category of the phoneme in the citation form of the word. Phonemic voicing is modeled as a 2 level factor: voiced, and voiceless, coded as 1 and -1, respectively.

- OBSTRUENT TYPE : A 2-level factor: stop, or fricative, coded as 1 and -1, respectively.

- CONSONANT DURATION: The duration of the final consonant as measured by corpus annotators. This is a continuous variable.

Both phonetic and phonemic voicing were included in the model because it was not known if there might be an effect of actual voicing above and beyond the effect of phonological voicing. We soon found that phonetic voicing did not differ appreciably from phonemic voicing, and it was dropped from the analyses. For the remainder of the paper, "voicing" will refer to phonological voicing. Although vowel duration based on vowel quality is not actually binary, data sparsity for certain low-frequency vowels makes using vowel quality itself problematic as a finer-grained determiner of inherent duration.

We use a measure of speaking rate that is based on the duration by which each word differs from the average of the mean values of its individual phonemes. This measure is based on similar metrics in which an expected duration is compared to an observed duration (e.g., Priva, 2010; Gahl et al., 2012; Gahl and Baayen, 2022). Such measures are used in order to make estimates of rate as independent of segment duration as possible. In principle, the two will be linked because it is not possible to avoid a confound when using acoustic data that has not been generated in explicitly rate-controlled contexts. Ambiguity in attributing duration differences among different factors is partially resolved by calculating deviation over the entire word.

Tanner et al. (2019) report a voicing effect for phrase-final monosyllabic words in the

Buckeye Corpus. Therefore, we attempt to reproduce their analysis as closely as possible. Nevetheless, there remain differences in how the samples were selected. Tanner et al. (2019) defined pre-pausal as preceding a silence of at least 150 ms, although they do not specify how they determined silent intervals. Our measure was more conservative, using only tokens with segmentation that indicated that the talker had stopped speaking long enough for the transcription to include an interval with silence, noise, or change in talker. This measure corresponds more closely to utterance-final position. Tanner et al. (2019) also excluded vowel tokens under 50 ms. We kept in tokens under 50 ms because that is not unusual for casual speech (there were 167 such tokens across stops and fricatives in utterance-final position). We also did not include stops that had been transcribed as flapped or glottalized, which further reduced the sample size, but gave us more homogeneous data. Tanner et al. (2019) included tokens with complex onsets and codas, while we analyzed only CVC forms. It is not clear whether they included reduced forms of polysyllabic words, or corrected for phonetic transcriptions of voiceless obstruents when the underlying value was voiced. Our final token count was 3200, while Tanner et al. (2019) analyzed 5500 tokens.

2.3 Methods

All statistics were performed using the lme4 package in R. Linear mixed effects models were run using the function lmer, fit by REML. T-tests used Satterthwaite's method, and the lmerTest function was used to obtain estimated p-values. All continuous numerical variables were log-transformed and mean-centered to approximate a normal distribution with a mean of zero. Following Tanner et al. (2019), we normalize by dividing by two standard deviations. Random intercepts for word and speaker were included in all models. Place of articulation of the final obstruent, although known to affect consonant duration, was too small of an effect to significantly improve model fit, and was therefore left out of the final model. Due to the asymmetric distribution of the data, it was not possible to used paired data in analyzing the voicing effect. All factors were sum-coded so that each individual factor was assessed at the mean value of all other factors. Three-way interactions were avoided for reasons of interpretability as well as model convergence.

2.4 Results

For each variable, the average value of its levels (if a factor), or of its range of values (if a continuous numerical variable) was the baseline for analysis. This allows us to conceptualize the results in a way that is similar to ANOVA, where each effect is an adjustment to the average value for the model. For example, the effect of Vowel Class is determined by whether the average duration of the class of Long vowels is significantly different from the global vowel duration average, calculated over both Long and Short vowels.

We begin the analysis with a simple model of vowel duration, containing no interactions, but using random intercepts and slopes. See Table 1. Separate models were run for phrase-final and phrase-medial tokens because we expected there to be a much larger effect for phrase-final tokens. Phrase-medial and phrase-final results, however, turned out to be very similar. We report the results of the phrase-final models below. All model formulae are included in Appendix C.

***** TABLE 1 ABOUT HERE*****

As expected, there was a significant main effect of speaking rate. Longer vowel durations were found at slower than average speaking rates. Word frequency also had the expected negative effect on vowel duration, such that words with higher than average frequency had shorter vowel durations. As predicted, long vowels were also longer than the average of long and short vowels, and high vowels were shorter than the average of high and low vowels. In this model, and also in the corresponding model of phrase-medial tokens, a small voicing effect is found. We used the vif function from the faraway R package to test for multilinearity: how strongly correlated a give model factor is with one more of the other model factors. A value of 1 indicates that there is no correlation. For the simple voicing model the vif scores are the following: Vowel height: 1.16; Vowel class: 1.09; Speaking rate: 1.23; Voicing: 1.30; Obstruent type: 1.19; Frequency: 1.19.Values less than 5 are generally assumed to be unproblematic (Gareth et al. 2013).

***** TABLE 2 ABOUT HERE*****

Our next step was to substitute consonant duration for voicing. See Table 2. In this second model the results are similar, except obstruent type is now significant: vowels are longer before fricatives than stops. There is also a significant negative effect of consonant duration on vowel duration. VIF scores for the model factors are: Vowel height: 1.11; Vowel class: 1.09; Speaking rate: 1.92; Consonant duration: 1.99; Obstruent type: 1.31; Frequency: 1.18.

Comparison of the two models using the anova function shows that the model with consonant duration provides a better fit to the data than the model with voicing. This is quantified by the difference in AIC score, which provides a measure of information loss, and is lower for the consonant duration model (1306) than the voicing model (2044).

Approaching the analysis from the other direction, we created a more complex model that was closer to that of Tanner et al. (2019). Their model includes interactions between voicing and frequency, voicing and vowel type, voicing and obstruent type, and voicing and word class. They included random intercepts for speaker, word, and vowel quality, as well as random slopes for speaker by frequency, vowel type, obstruent type, word class, and by the interaction of voicing and obstruent type. Random slopes were also included for word by both speaking rate measures that they used. This model overfits our data and does not converge. We had to modify the model significantly for this reason, as well as to make the voicing and duration models comparable. See Appendix (C) for the full details of this analysis. In the complex voicing model voicing did not have a significant effect on vowel duration. However, certain interactions reached significance. The effect of voicing showed an increase at slower rates, and a decrease for higher frequency words. In the duration model consonant duration was significantly negatively correlated with vowel duration. No interactions reached significance. Model comparison once again shows that the duration model provides a better fit to the data.

2.5 Obstruent duration

In laboratory settings, voiced obstruents are consistently found to be shorter than voiceless obstruents (e.g., Klatt, 1976; Umeda, 1975; Miller and Volaitis, 1989; Chen, 1970; Luce and Charles-Luce, 1985). In conversational speech, however, the duration distributions are almost completely overlapped, as can be seen in Figure 2, which shows the raw consonant durations as both counts and probability densities. These distributions look strikingly similar to the vowel duration distributions from Figure 1.

*****FIGURE 2 ABOUT HERE***

A regression model with obstruent duration as the dependent variable shows that voicing is significantly negatively correlated for the phrase-final data. See Table 3. No interactions reached significance. Although these results, in and of themselves, cannot prove the hypothesis that vowel duration differences arise from consonant duration differences, they are consistent with that hypothesis.

****TABLE 3 ABOUT HERE***

2.6 Summary & Discussion Of Corpus Results

The corpus results show that, while voicing is a significant predictor in the simple model, consonant duration performs even better. Voicing, however, is not significant in the complex model, failing to replicate the finding of Tanner et al. (2019). We also see a dependence of the voicing factor on absolute durations in interaction terms with speaking rate and frequency. This is expected if the voicing effect is actually an effect of consonant duration, where the difference in obstruent duration between voiced and voiceless obstruents increases with increasing duration.

On the other hand, laboratory production studies find that there is a large and consistent voicing effect; pre-voiced vowels can be up to 50% longer than pre-voiceless vowels (Peterson and Lehiste, 1960; Mack, 1982; House, 1961; Luce and Charles-Luce, 1985; Umeda, 1975). Similarly, voiceless stop closure durations can be from 25% to 50% longer than voiced stop closures (Chen, 1970; Luce, 1986). These discrepancies can be explained by the large difference in absolute durations between the corpus and the laboratory. Vowel durations in those studies were reported in the range of 175 to 300 milliseconds (Peterson and Lehiste, 1960; Mack, 1982; House, 1961; Luce and Charles-Luce, 1985; Umeda, 1975), with voiceless stop closures ranging from 95-140 milliseconds (Luce and Charles-Luce, 1985; Chen, 1970). For the vowel tokens in the Buckeye Corpus, on the other hand, durations this long are rare. Among the set of CVC words ending in voiced obstruents, less than 7% reach durations of 200 ms or above. Even restricting the sample to just characteristically longer vowels, only 13% of such tokens fall in this range. Median vowel duration over the complete set of CVC words used in this study is only 83 ms. Median vowel duration for just the voiced tokens is actually lower than that, at 75 ms. Similarly, only 9% of CVC-final voiceless stops reach durations of 100 ms or above in the Buckeye Corpus, while the median closure duration is 46 milliseconds.

3 A Production Study

We take the corpus results, in conjunction with the production literature as a whole, to provide strong preliminary support for the Expandability Hypothesis. However, because paired data are not available in the corpus,[7] our predictions must be confirmed in a setting where sources of variation can be controlled for. In this section we report the results of a production experiment in which we asked native English speakers to repeat a series of CVC minimal pairs at varying rates. This allows us to directly compare the lengthening behavior of final voiced obstruents to that of final voiceless obstruents. Additionally, the difference in obstruent duration can be compared to the difference in vowel duration as a function of speaking rate.

3.1 Methodology

3.1.1 Participants

All participants were undergraduate students at The Ohio State University who were given course credit for completing the experiment. A total of 45 participants were run: 25 were female,

[7] And may not be readily available to listeners in any case, if the distributional properties are similar across other spoken corpora.

and 20 were male. The average age of participants was 21. Of this group, 11 were excluded from analysis for the following reasons: they reported hearing issues (3); they reported learning a language other than English before the age of 7 (7); they did not learn English until after the age of 7 (1).

3.1.2 Procedure

Participants were seated in front of a computer monitor inside a sound-attenuated booth. Continuous audio was recorded from a desktop microphone using the sound editing software Audacity.[8] Participants were instructed that they would be asked to speak into the microphone in response to prompts on the computer screen. The entire experiment took less than an hour to complete.

The experiment began with a practice block to acclimate participants to the experimental task, and the different repetition rates involved. Prior to the start of the practice block, participants were given the following instructions:

> *A + sign will appear on the screen. It will be black to begin with, then will change to red, and keep alternating. Your job is to repeat the word on the screen every time + changes color. Try to use the entire time that the + does NOT change color to say the word. Keep going until the flashing stops. Press any key when you are ready to practice with the word "lab".*

For the first trial, participants saw the following text: *"Here's the fastest speed"*. The word "lab" appeared 1.5 seconds later. The word stayed on the screen as the "+" immediately appeared and began to change color. Color changes occurred 8 times. At the end of the 8 cycles, a new trial began. For each new trial, participants were alerted to the change with the following text: *"A little slower"*. The same word then appeared 2 seconds later. There were 5 different rates, corresponding to the time it took for the plus sign to change from black to red: 350, 550, 750,

[8] Available at http://audacity.sourceforge.net.

950, and 1150 ms. The slowest and fastest rates were chosen to be as extreme as possible while still being within the ability of participants to match.[9]

At the end of the practice session participants were told that they could begin the experiment whenever they were ready. The experimental trials were identical to the practice except that the rates went in order from slowest to fastest. Participants were presented with the following text: *"You will begin with the SLOWEST speed, and the flashing will become faster"*. Subsequently, each rate change was signaled with: *"The speaking rate will now speed up a bit"*. Trials were blocked by word, such that participants experienced all rates before beginning with a new word. At the end of a given block, participants were alerted that *"The next item will now appear on the screen",* with a pause of 2 seconds before the word appeared. Word order was randomized across participants, but the order of rate presentation was fixed. Each word/rate pair was presented once.

The minimal pairs reported in this paper were chosen to vary across vowel quality (o or i), consonant manner (stop or fricative), and final consonant place (coronal or labial): feet/feed, thief/thieve, lobe/lope, and doze/dose. Differences in part of speech and morphological complexity were largely unavoidable in constructing CVC minimal pairs, but those factors are not expected to show interactions with speaking rate. No effort was made to balance word frequency, beyond the avoidance of archaic forms, for the same reason. While higher frequency words would be expected to be somewhat shorter across the board, there was no reason to believe that speaking rate would affect the individual segments differently.

3.2 Data Selection and Annotation

Each participant produced approximately 8 tokens of each word at each rate. To avoid edge effects, and fluctuations in rate, a single representative token from the center of the group was selected and measured. Because each token was surrounded by other tokens at the same

---

[9] Note that the fastest change time, 350 ms, is quite long in terms of vowel duration alone, as measured in the Buckeye Corpus. This presumably reflects the fact that coarticulation and reduction, along with prosodic organization, allow for individual segments to be much shorter in normal speech than in a laboratory word-repetition task.

repetition rate, it was possible to segment both the closure and the release interval for each stop. Occasionally, at the fastest rates, final stops did not have a clear release. In those cases, the end of the stop was set to the end of the voicing bar (for voiced stops), or the point at which the amplitude returned to background levels (for the voiceless stop). Background level was estimated by the amount of noise visible during the gaps between successive words.

The most ambiguous cases involved the segmentation of the sonorant /l/ from the following /o/ vowel, given a large degree of coarticulation. At faster rates, the point at which the release of the final /d/ became the initial fricative of the following token of "feed" could also be hard to determine. This was also true of the final /v/ and the initial /θ/ in "thieve" sequences. Measurement variability is likely to be highest in those contexts. Note, however, that any measurement errors for these kinds of tokens will only introduce error for one of the measured variables. For "lo", the vowel duration will be affected by where the segment boundary is placed, but not the final p/b. For "<u>feed</u>feed" and "<u>thieve</u>thieve" the coda duration will be affected by where the segment boundary is placed, but not the vowel duration. Furthermore, any possible annotator bias in segmenting the sequence "lo", for example, would have a minimal impact on the results, firstly because each participant was assigned to a single annotator, meaning that any effect could be absorbed into a random effect by speaker, and secondly, because both the voiced and voiceless minimal pairs would be segmented in the same way, such that the voicing effect (the difference in vowel durations) would not be affected by any bias. There is a possibility of resyllabification for the fastest word repetition rates, but this is only likely for p#l and b#l sequences in the lope/lobe pair. If such resyllabification occurred, we might expect the stop to be shorter, with reduced aspiration. In fact, this might explain the abrupt drop in VOT between rates 4 and 5 for this word pair (See Fig. 3).

The data for the first two word pairs (feet/feed, thief/thieve) were randomly assigned to three undergraduate research assistants for annotation. One of the authors and two of the RAs then re-measured a subset of the data produced by the other two annotators. Discrepancies between any two raters were discussed as a group to establish shared criteria for ambiguous tokens. The two RAs then individually reviewed their previous measurements and made adjustments where their original segmentation did not meet the discussed criteria. The same two RAs each also re-measured half the data of the third RA who had left the lab at that point. The second set of words (lobe/lope, doze/dose) were measured later, by an additional two RAs. Measurement verification was conducted in the same way. It was stressed that the most important criterion was consistency. As a final check, 2% of all tokens from each annotator were re-measured by the first author, selected in pairs in order to assess the discrepancy in the measured voicing effect. In terms of absolute durations, vowel measurements differed by an average of 12 ms, and total consonant durations difference by an average of 21 ms. The difference in vowel duration between voiced and voiceless minimal pairs differed by 13 ms, and the difference in total consonant duration by 30 ms. However, because the durations were sometimes longer than the first author's measurements, and sometimes shorter, the actual effect of discrepancies in this sample of data were much smaller: 6 ms shorter for vowel duration; 12 ms shorter for total consonant duration, a vowel duration difference that was 3.6 ms smaller, and a consonant duration difference that was 15 ms larger.

Occasionally the voiced stops and fricatives at the slower repetition rates were produced with a final epenthetic schwa. There were 29 such tokens. Any words with final schwa were removed from the analysis. In many cases, participants produced dose and doze tokens that were difficult to disambiguate. Two such participants were removed due to their productions of final s and z being practically identical. Five participants were removed for either failing to vary their speaking rate significantly across trials, or varying only inter-word pause duration rather than word duration. An additional participant was removed due to adopting a sing-song (high-low) prosody to the word repetition. The results from the 26 remaining participants, and the 5049 measured tokens, are given below. Praat (Boersma and Weenink, 2009) was used for segmentation and annotation.

3.3 Results

In Fig. 3 final consonant durations for the stop-final words are plotted as a function of repetition rate (shown as a number between 1 and 5, where 5 is the fastest rate, and 1 the slowest). Voiced and voiceless tokens are plotted separately, and three different duration measures are given: closure (black), VOT (light gray), and the sum of the two (TDur: dark gray). Closure duration for final voiced stops varied relatively little across repetition rates. However, most stops were also produced with a period of aspiration (VOT). Voiced stops show a clear increase in total duration as rate decreases, but one that appears to plateau at the slowest rates. For voiceless stops, closure duration increases steadily, patterning very closely with VOT. Because both duration measures show dependence on rate, total duration was used as the dependent variable for testing the Expandability Hypothesis.

FIGURE 3 ABOUT HERE

Figure 4 provides duration data for the full set of words, both vowel duration (triangles), and total obstruent duration (filled circles). Visual inspection shows that larger vowel durations were reached by the voiced member of each minimal pair, while larger obstruent durations were reached by the voiceless member. There is also a larger difference between consonant and vowel durations for voiced-final tokens across all repetition rates, and that difference increases with decreasing repetition rate.

FIGURE 4 ABOUT HERE

A linear mixed-effects model was fit to the vowel duration data as a function of repetition rate and consonant duration. Consonant duration was treated as a continuous variable, and repetition rate, as an ordinal variable. Random intercepts for participant and word were included. Random slopes were not used as they caused the model to fail to converge, or led to singularity. As expected, a significant (linear and quadratic) effect of speaking rate was found (vowels were longer at slower speaking rates). There was also a main effect of consonant duration; vowels were longer when the coda consonant was shorter. The interaction between rate and consonant duration also reached significance; the negative effect of consonant duration was strongest at the slowest rates. See Table 4. Only significant effects are shown.

TABLE 4 ABOUT HERE

A separate model of vowel duration as a function of rate and voicing behaves very similarly. As before, random intercepts were used for participant and word, but random slopes were not used as they caused the model to fail to converge, or led to singularity. Main effects of (linear) rate and voicing are found (reference level is Voiceless), as well as an interaction between voicing and rate such that the positive effect of voicing is strongest at slower rates. Model comparison shows that the consonant duration model only does slightly better in fitting to the data (AIC score of 11972 vs. 12018) due to the fact that duration and voicing are highly correlated in the production data.

The model of consonant duration as a function of voicing confirms the interpretation that the "voicing" effect is driven by consonant duration. See Table 5. Random intercepts for participant and word were included. Random slopes were not used as they caused the model to fail to converge, or led to singularity. A fully crossed rate, voicing, and manner model produced significant main effects of speaking rate (linear), voicing, and manner. Fricatives were significantly longer than stops (reference level is Stops). A significant interaction between rate (linear) and voicing was also found, indicating, as expected, that differences in duration between voiced and voiceless consonants increased with decreasing repetition rate. An interaction between manner and rate (linear) also reached significance: the difference in duration between fricatives and stops was even larger at slower rates. Only significant interactions are shown.

****TABLE 5 ABOUT HERE****

A final analysis of the paired duration differences confirms the negative correlation between the *difference* in duration of voiceless and voiced consonants, and the *difference* in duration of their preceding vowels. Random intercepts for participant and word were included. Random slopes were excluded as they caused the model to fail to converge, or led to singularity. Adding manner to the model also resulted in singularity. The final model of ΔV ($= V_{VL} - V_{VD}$) included rate and consonant duration difference ($\Delta C = C_{VL} - C_{VD}$) and their interactions. A significant main effect of rate (quadratic) and ΔC were found, and significant interactions between ΔC and rate, for both the linear and the quadratic terms. Thus the voicing effect (ΔV) is shown to be larger for larger negative values of ΔC, which are enhanced at the slowest speeds. See Table 6. Only significant factors are shown.

****TABLE 6 ABOUT HERE****

3.4 Discussion

These results strongly support the Expandability Hypothesis. Firstly, we confirm the predicted difference in lengthening between voiced and voiceless consonants in coda position, paralleling what has been repeatedly found for consonants in initial and medial position (Port, 1976, 1981; Miller and Baer, 1983; Miller and Volaitis, 1989; Volaitis and Miller, 1992). There is a difference in consonant durations at all rates,[10]and there is also a large difference in the slopes of the duration curves. The difference in consonant duration increases with decreasing rate, as does the vowel duration difference. Pairing consonant duration differences with vowel duration differences at each speaking rate shows that the strength of the voicing effect is significantly correlated with the size of the consonant duration difference. The significant interaction between rate and consonant duration (vowels), and between rate and voicing (consonants), is precisely what is predicted if vowel duration differences derive from consonant duration differences, rather than depending on phonetic voicing, or an abstract phonological voicing feature. In fact, absolute vowel duration differences and consonant duration differences are very close. Rhyme (VC) durations were significantly different between voiceless and voiced, but not large, at 26 ms. For

[10] Note that the shortest vowel durations in this study are between 150 and 200 ms, already in the upper range of values found in the conversational speech of the Buckeye Corpus.

final stops, the rhyme duration difference was only 2.8 ms. These results probably over-estimate the degree to which vowel and consonant duration are traded off, given that the experimental task is highly unnatural, and likely to bias more towards uniform syllable duration than natural speech contexts.

4 The Expandability Hypothesis: Modeling the Corpus Data

The corpus and production study results, combined with the previous research summarized in Section 1, strongly suggest that final obstruent duration trades off against preceding vowel duration, and that the size of the resulting voicing effect depends on absolute duration. To account for both of these properties, in addition to the fact that apparent compensation is not "perfect" (cf., Chen, 1970; Keating, 1985; Port and Dalby, 1982), we propose a competition-based model where trade-offs in duration arise, not from isochrony, but from pressures to meet certain duration targets, none of which can be fully satisfied.

4.1 A Competing Constraints Model of the Voicing Effect

In this section we model the voicing effect as the outcome of a competition between conflicting duration targets at the segment and syllable level. Each segment possesses an inherent elasticity which is implemented as the weighting factor on a constraint that acts to keep the segment at its preferred duration. All constraints are gradiently "violable", such that no given constraint need be perfectly satisfied (allowing for undershoot and overshoot).

Constraints, in turn, are implemented as Normal probability distributions. Each distribution assigns the highest probability to its preferred duration (the mean of the distribution), and smoothly decreasing probabilities for durations both longer and shorter than that mean. The variance of the distribution controls how quickly the probability decreases.[11] The smaller the variance, the more rapid the decrease, and the greater the resistance to deviations from the mean. Its variance thus acts effectively as a weighting factor for each constraint. This means that, all else

[11] Note that distribution variance is not a measure of actual duration variance. The latter is determined by the interaction of all constraints.

being equal, a segment with a broader probability distribution will be lengthened or shortened more than a segment with a narrower probability distribution. Variance thus also maps to segment elasticity. Constraint "competition" in this model is realized through maximization of the joint probability function over all constraints. This function exhibits the desired behavior: one constraint may be "violated" to a greater degree (decrease in probability) if this allows another, more highly weighted constraint to be less "violated" (*greater* increase in probability).

The results reported here are for VC syllables. See Appendix (B) for the treatment of CV syllables, and the "voicing" effect in onset position. The three segment-level constraints for the voicing effect model are shown graphically in Fig. 5. Voiced and voiceless obstruent constraints are given the same mean value in these simulations, differing only in their variance. Target syllable duration is treated as a random variable, an external specification for a specific speaking rate or duration. Thus, each pair of values, (V, C), derived by the model are conditioned on a particular target syllable duration. In addition, two inter-segmental constraints specify preferred values for the $\frac{C}{V}$ duration ratio and the $\frac{V}{\sigma}$ duration ratio, respectively. The $\frac{V}{\sigma}$ duration constraint forces vowel duration to lengthen with lengthening target syllable duration, while the $\frac{C}{V}$ duration constraint requires consonant duration to do the same. Together they enforce monotonic behavior for both segments, such that they never shorten with an increase in target syllable duration, or lengthen with a decrease in target duration.

FIGURE 5 ABOUT HERE

If target syllable duration is strictly enforced, then this model will produce perfect compensation, meaning that a given voiced syllable and a given voiceless syllable can always be set to the same duration. To avoid this, we introduce a final, violable, constraint for matching the target syllable duration (σ_T). This constraint has a probability distribution centered at zero, over the variable $\frac{\sigma_T - \sigma}{\sigma_T}$: the normalized difference between actual and target durations. In the absence of an imposed target duration, all segments would be realized at their preferred absolute durations. In actuality, however, duration will always depend on how quickly one is speaking. In the model, this is the result of the competing pressure to reach the target syllable duration and the resistance of the individual segments to expansion or compression.

For a given target syllable duration, the model conducts a brute force search for the durations of the coda consonant (D or T) and vowel (V) that result in the highest joint probability over the entire set of constraints.[12] Although the model simply tries all possible combinations of values, the search space is restricted within a range where the maximum possible vowel duration is set to the maximum syllable duration, and the minimum vowel duration is set to half the maximum syllable duration (such that consonant duration can never be larger than vowel duration). A fixed step size of 1 ms for both consonant and vowel is used to search this space. Each variable is assumed to be independent, therefore the joint probability is given as the product of the individual probabilities. See Appendix (B) for further details of the model.

Figure 6 shows the result of running the model for the set of target syllable durations ranging from 30 to 700 ms, sampled at 20 ms intervals (x-axis). On the y-axis, vowel duration, consonant duration, and syllable duration (V+C) are plotted for both voiced (black) and voiceless (gray) syllables. Each point on the graph corresponds to a VC word produced at the specified duration/rate. For example, for a target syllable duration of 330 ms, and a voiced-final syllable, the optimal vowel duration is 249 ms, and the optimal voiced obstruent duration is 62 ms. These points are shown as filled circles in Figure 5. Note that actual syllable duration is less than the target, at 311 ms. For a voiceless-final syllable, on the other hand, the optimal vowel duration is 238 ms, and the optimal voiceless obstruent duration is 80 ms (for a syllable duration of 318 ms). These points are shown as unfilled circles in Figure 5. The vowel, like each obstruent type, prefers the mean duration of its probability distribution (100 ms in this case). It is forced to lengthen due to the pressures of the other constraints. Because the voiced obstruent constraint is more highly weighted than the voiceless obstruent constraint (smaller variance), it does not shift

[12] Following Browman and Goldstein (1986) inter alia, we assume that there is a preferred timing relationship for a VC syllable which governs the degree of overlap between the articulatory gestures corresponding to the nucleus, and those corresponding to the coda. This parameter affects the apparent acoustic duration of the vowel, i.e., the portion that is not masked by the following consonant. Although we assume that modifications to this phasing relationship are possible, it does not vary in the current model. In all cases, there is no overlap between the two segments, such that the acoustic syllable duration is given by the sum of the vowel and consonant durations.

as far from its preferred duration (at 50 ms). Therefore, the vowel is forced to lengthen more when it co-occurs with a voiced obstruent than with a voiceless obstruent.

At shorter target syllable durations, voiced and voiceless obstruents (black and gray solid lines, respectively) are more or less identical in duration; preceding vowel durations (black and gray dashed lines) are also identical within the same range. As target syllable duration continues to increase, the consonant durations start to diverge. Because of its much smaller variance, the voiced obstruent not only resists lengthening more strongly than the voiceless, but that resistance also grows faster, leading to smaller and smaller increases in duration. As a result, either the vowel must lengthen more, or the divergence from the target syllable duration must increase, or both. As the obstruent duration difference continues to increase, the vowel duration difference will also continue to increase, meaning that the magnitude of the voicing effect will increase with increasing duration.

FIGURE 6 ABOUT HERE

As we have seen, because the constraint to match target duration competes with other constraints, a given syllable does not always match the target exactly. Thus this model allows for under- and over- compensation. This occurs only when imperfect matching would increase over-all probability. In the expansion regime, where vowel and consonant durations both increase past their preferred means (indicated by the filled circles in Fig. 6), both voiced and voiceless syllables are shorter than they would be if target syllable duration were strictly enforced, and the degree of under-compensation increases with increasing duration. This is the difference between the gray line ($\sigma = \sigma_T$), and the two dotted lines that indicate actual syllable duration in Fig. 6. The increasing deviation from target is partially due to the fact that variance is expressed as a proportion: a larger deviation is tolerated for a longer syllable.

Voiced syllables are also systematically shorter than voiceless.[13] This is because even the highly-expandable vowel has a preference for its mean duration. A balance is struck between the length of the vowel and the amount of deviation from the target. In the compression regime,

[13] This is not necessarily the case under different parameter values.

where segments shorten past their preferred durations, syllable durations are slightly longer than the target. For consonant durations below the mean, voiceless obstruents become slightly shorter than voiced. This occurs because elasticity is bi-directional; voiceless obstruents are both more expandable and more compressible than voiced stops.

Using this model, we simulated the corpus data by sampling 1000 points from a Normal distribution of target syllable durations,[14] durations that fall mostly in the range where there is a negligible difference in consonant duration. This sample is depicted by the light blue vertical bars in Fig. 6. Each point from the sample represents a VC word produced at a given target duration/rate. For each of these 1000 points, the vowel and consonant durations that optimized joint probability were calculated, for voiced and voiceless syllables separately. Because words within each distribution were treated as identical (VD or VT), all words with the same target syllable duration had the same segment durations. The resulting data are plotted as probability density functions in Figure 7, for comparison with the corpus data in figures 1 and 2. Since only one type of vowel (fixed variance) was used, the simulation results do not entirely map to the corpus. However, they demonstrate that vowels of different inherent length are not necessary to derive the observed duration dependence of the voicing effect.

FIGURE 7 ABOUT HERE

As a proof of concept, the model does quite well at capturing the critical behaviors that motivated our re-analysis of the voicing effect in English, and without a directly compensatory mechanism. Languages other than English can be modeled by changing the relative variances of the obstruent probability distributions. A smaller difference in elasticity leads to a smaller voicing effect. The model can also capture the interaction between the voicing effect and vowel quality, using a lower elasticity parameter for inherently shorter vowels. Reducing the variance of the vowel probability distribution, but keeping all other parameters the same, results in a smaller

[14] All durations less than 30 ms were set to 0.

voicing effect, and shorter syllables over-all. The voiced obstruents become slightly longer under these conditions, but the largest change is in how closely the target syllable duration is approximated. In this model, the reduced variance of the vowel results in shorter vowel durations, causing greater undershoot at the syllable level. Qualitatively, this behavior is consistent with the finding that the voicing effect is significantly reduced in preceding vowels that are inherently short (Umeda, 1975; Crystal and House, 1982; De Jong, 2004). Note that the difference in duration between the obstruents themselves can, in principle, still grow quite large. See Appendix (B). Because very few studies on the voicing effect report final obstruent durations, it remains to be seen whether this prediction is borne out.

Our model can also be used to model the behavior of CV syllables that differ with respect to the voicing of an onset consonant. In this model it is necessary to make the assumption that it is actually the rhyme, rather than the entire syllable that is the relevant prosodic unit. Additionally, a different phasing relationship is used to capture the onset-nucleus timing: one in which the consonant completely overlaps with the vowel (see, e.g., Browman and Goldstein, 1988). Because target rhyme duration does not affect the onset consonant, there is no trade-off of consonant duration with vowel duration. However, longer onset durations mask more of the vowel. As the durations of the voiced and voiceless obstruents diverge, vowels phased with voiceless obstruents become acoustically shorter than vowels phased with voiced obstruents. The behavior of this model is very similar to the short vowel model: a smaller voicing effect, but a similarly large divergence in obstruent duration. These two features are consistent with the known data. Voiced obstruents lengthen very little in speaking rate studies of CV syllables, while voiceless obstruents lengthen considerably, and vowel duration differences, when observed, are quite small (Allen and Miller, 1999; Pind, 1995; Miller and Baer, 1983; Miller et al., 1986). See Appendix (B) for more details of this model.

The competing constraints model does not differentiate between sources of lengthening, modeling only what occurs at the segment level to meet specified targets at some higher prosodic level, whether rhyme, syllable, word or foot. For very slow speaking rates, of the kind encountered in laboratory speech, a robust voicing effect can be observed (e.g., Peterson and Lehiste, 1960; Mack, 1982; House, 1961; Luce and Charles-Luce, 1985; Umeda, 1975). Similarly, pre-pausal lengthening can also produce a significant voicing effect. A particularly large final lengthening effect in English (e.g., Delattre, 1966), we conjecture, may be largely responsible for the particularly large voicing effect in this language.

5 Further Tests of The Expandability Hypothesis

In the previous sections we have shown that vowel duration is better predicted by coda duration than by coda voicing. The implication being that the correlation between obstruent duration and voicing is the source of the apparent voicing effect. It has also been demonstrated that a model of competing durational constraints can qualitatively capture the duration trade-offs between consonant and vowel duration. However, the Expandability Hypothesis, in and of itself, does not explain the ability of listeners to reliably use vowel duration to predict post-vocalic obstruent voicing. In this section we will show that not only is the Expandability Hypothesis consistent with the perception literature, it is confirmed by certain results. For the remainder of the paper we will focus on word-final stops because there are often very limited cues to stops in final position, and it is primarily for stops that preceding vowel duration has been characterized as a contrastive cue.

5.1 Perception of voicing in final position

A review of the perception literature in Section 1.6 has shown that other cues to the voicing contrast are likely to be stronger than preceding vowel duration, and categorical perception results may only be possible with highly impoverished stimuli. Meanwhile, categorical perception results have been obtained by varying obstruent duration alone (e.g., Denes, 1955). Based on these results, we hypothesize that listeners are using stop duration itself as the cue to voicing when final stops are both voiceless and unaspirated. Vowel duration factors into the classification decision insofar as it provides information about stop duration indirectly, as a measure of speaking rate.[15] In essence, the listener's task is to decide whether what they are hearing is a voiced stop spoken slowly or a voiceless stop spoken quickly. Shorter vowel durations, which comprise the majority of the corpus data, correspond to speaking rates at which voiced and voiceless stop durations are not significantly different from one another. In this range, vowel duration is ineffective as a cue to voicing. Only as speaking rate slows to the point where the voiced and voiceless expansion trajectories begin to diverge, does vowel duration become predictive.

The competition model of Section 4 is used to illustrate this hypothesis. See Fig. 8. The duration of the voiceless stop (gray solid line) gradually diverges from the duration of its voiced counterpart (black solid line), as the syllable is lengthened. This divergence is mirrored in the preceding vowel duration (gray dashed line – preceding voiceless stop; black dashed line – preceding voiced stop). If the listener is exposed to a relatively short vowel (Fig. 7a: upper horizontal dotted line), their expectation for the duration of the upcoming stop will be roughly the same regardless of whether it is voiced or voiceless (vertical difference between the lower open circles). An observed stop duration (lower dotted line) that falls close enough to both expected values is assumed to be acceptable for either member of the pair, and will not be sufficient to distinguish between the two in the absence of other cues.

FIGURE 8 ABOUT HERE

For a longer vowel, on the other hand, there is a larger difference in the expected durations of the voiced and voiceless stops. See Figure 7b. The same observed stop duration (lower dotted line) now falls significantly below both expected values. In a two-alternative forced choice task we predict that this stimulus should sound more like a voiced than a voiceless stop. In general, an ambiguous final stop of fixed duration should sound more and more like a voiced stop as vowel duration increases (speaking rate decreases). We assume that the category cross-over point from

[15] It is common practice to use stressed vowel duration as a proxy for local speaking rate (e.g., Crystal and House, 1982; Summerfield, 1981; Port and Dalby, 1982).

voiceless to voiced falls where the stimulus is significantly shorter than expected for a voiceless stop at that rate. After that, the likelihood of a voiced stop continues to increase (cf. Massaro and Cohen, 1983).

The foregoing can thus explain the increase in voiced responses with increasing vowel duration. However, given that we hypothesize that shorter vowels should not provide any cues to the voicing contrast, we would expect, all else being equal, that listeners would be at chance in identifying tokens in the short half of the continuum. Here it is the nature of the actual experimental stimuli that may bias perception strongly towards the voiceless stop. In the first place, ambiguous tokens are, by definition, phonetically voiceless. Depending on how exactly such stimuli were created, they may retain other cues to the original speech token from which they were generated, such as an F1 offset that is more consistent with a voiceless, than a voiced, stop. The synthetic stimuli used in Denes (1955), for example, were based on originally voiceless tokens. Whereas Repp and Williams (1985), using naturally produced stimuli, found a large perceptual difference between continua generated from an originally voiced stop (lab), versus an originally voiceless stop (lap). Voiced responses were about 40% higher for the former across all but the two longest vowel durations.

We therefore posit that the categorical perception results are due, firstly, to a default voiceless percept, based on residual cues that are more consistent with the voiceless member of the contrast, and secondly, to unusually long vowel durations. At the longest vowel durations (vanishingly rare in the speech corpus), we posit that the expected duration of a voiceless stop is so long that its likelihood approaches zero. For such extreme tokens, selection/perception of the voiced alternative may occur prior to actually hearing the final segment. However, it appears that the addition of a period of strong aspiration at the end of the stop is sufficient to switch the percept to voiceless.[16] Listeners may also be able to reliably select the voiced member of a minimal pair when final stops are entirely removed. We suspect that this is only possible in an explicit comparison task where listeners must label one token as voiced, and one as voiceless. In

[16] This was established anecdotally when the spliced stimuli were played for various audiences.

such a a task it is likely that listeners assume a uniform speech rate, leading them to attribute a somewhat longer vowel duration to the effect of a following voiced stop.

Additional support for this account of voicing perception comes from studies of the voicing contrast in initial position. It has been consistently found that the perceptual VOT boundary is longer than the boundary estimated from production data (e.g., Miller et al., 1986; Miller and Volaitis, 1989; Volaitis and Miller, 1992). However, the two boundaries coincide when naturally produced, unedited stimuli are used in the perception task. Nagao and de Jong (2007) suggest that the mismatch may arise from the fact that the stimuli typically used in perception experiments are artificially impoverished. In other words, the edited tokens are so ambiguous that they can only be confidently classified at very long VOT, or very slow speaking rates. The consistency in the reported perceptual cross-over point across experiments on word-final stops may be explained by the same artificiality. For voiceless closures with no audible release, the duration of the coda stop is indeterminate. Listeners may therefore assume a duration that is plausible given their language experience and consistent with experimental variables such as the inter-stimulus interval. It is therefore likely to be relatively stable across experiments involving native speakers of English.

5.2 Predictions

Our explanation of the perception results generates at least one testable hypothesis. We predict that a change in the perception of voicing should lead to a change in the perception of speaking rate. During the course of vowel production, it is assumed that a hypothesis about both speaking rate and following segment duration is generated by the listener. In the absence of any information about the duration of the following stop (silent and unreleased), we posit that listeners will infer a duration that is consistent with those hypotheses. For a particularly long vowel, an expectation for a following phonologically voiced stop should lead listeners to infer the expected duration for a voiced obstruent, and the speaking rate associated with that duration (as depicted in Figure 7b: the intercepts of the leftmost vertical line with the voiced obstruent duration curve and the x-axis, respectively). However, if listeners subsequently experience unambiguous release or aspiration cues, then we hypothesize that there should be a noticeable correction to both the perceived stop class and the perceived speaking rate. The voiceless stop should indicate that the speaking rate is actually slower than previously supposed (represented by the x-intercept of the rightmost vertical line in Figure 7b).[17] Sanker (2019) has shown that the judgment of whether a vowel is "long" or "short" depends not only on the duration of the vowel, but on whether it is followed by a voiced or a voiceless obstruent. For vowels preceding voiced obstruents, longer durations are required to elicit a "long" response. Although she did not report obstruent duration, we interpret her results as deriving from the expectation for a specific vowel duration given the unambiguous obstruent duration and its voicing. Vowels shorter than this expected value would be perceived as "short", and vowels longer than this value would be perceived as "long".

The Expandability Hypothesis also predicts that it should be possible to find apparent compensation with segments other than immediately preceding or following vowels, as long as they are more expandable than voiced obstruents. This is corroborated to a certain extent. A difference in nasal duration preceding voiced versus voiceless stops has been found both for monosyllabic words of the form "dens/dense" (Raphael et al., 1975; Port and Cummins, 1992; Beddor, 2009), and polysyllabic words of the form "cantor/candor" (Vatikiotis-Bateson, 1984). Furthermore, Raphael et al. (1975) find that both vowel and nasal duration affect perception of voicing on final stops. In an eye-tracking study by Beddor et al. (2013), participants heard CVND words (such as "bend"), CVNT words (such as "bent"), and $C\tilde{V}C$ words ([bɛ̃d] vs [bɛ̃t]), in which the nasal was missing but the vowel was nasalized. They found that, for $C\tilde{V}C$ tokens, participants were overall more likely to fixate on the image corresponding to the CVNT word than the CVND word. They interpret this result as deriving from listener expectation that the nasal gesture will be coordinated differently in the two contexts: initiating earlier before a voiceless stop, and later

[17] The expected voiceless obstruent duration for that vowel duration is also expected to be longer. However, because speaking rate perception likely depends more on vowel duration than consonant duration, a change in the percept of voicing alone, without a change in the actual obstruent duration may be sufficient to trigger a change in the perception of speaking rate.

before a voiced stop. However, no explanation is offered as to why the phasing relationship should be different in the two contexts. This difference, however, can be accounted for under the Expandability Hypothesis if the competition at the word (or syllable) level affects both the duration of individual gestures, as well as their phasing, as occurs under changes in speaking rate (e.g., Stetson, 1928; Hardcastle, 1985), and other types of prosodic lengthening (e.g., Byrd and Saltzman, 1998; Byrd et al., 2000). A shorter voiced stop would thus correlate with both longer tautosyllabic segments, as well as a preceding VN sequence that is less coarticulated. Less coarticulation, in turn, would result in less vowel nasalization. Thus, a highly nasalized vowel is more likely to occur preceding a [t] than a [d].

An additional corollary of our account of the voicing effect is that actual voicing, or any feature other than length, is not required for a "voicing" effect to arise. In fact, active phonetic voicing cannot be a requirement when the strongest effect is seen in English pre-pausally, where final voiced obstruents are likely to undergo devoicing. Given our hypothesis, however, it should be possible to find a "voicing" effect involving segments that have low elasticity for a reason not related to historic voicing. Beguš (2017) finds that stop duration correlates negatively with preceding vowel duration not just for voiced and voiceless stops in Georgian, but for ejectives as well, with ejectives intermediate between voiced and voiceless stops in terms of both consonant duration and preceding vowel duration.

In principle, any apparent temporal compensation phenomenon could be modeled using the competing constraints framework (see Section 4.1). All else being equal, we might also predict that an appreciable difference in consonant duration should lead to a complementary difference in preceding vowel duration in monosyllabic words. However, it may prove difficult to isolate elasticity-based effects from other factors that affect syllable duration. For example, vowels in monosyllables closed by nasals have been found to be as long, or longer, than vowels in monosyllables closed by voiced obstruents in English (e.g., Peterson and Lehiste, 1960; Umeda, 1975; Crystal and House, 1988; House and Fairbanks, 1953), which is the opposite of what one would expect for a sonorous segment like a nasal. However, both Crystal and House (1988) and

Klatt (1975) report nasal durations that are comparable to those for voiced stops. Thus, it may be the case that nasals (and possibly other sonorants) are not as elastic as might have been expected. Another possibility is that the phasing relationship between vowel and coda may be different in the case where the two gestures can overlap significantly without masking. Thus vowels may be measured as longer, and nasals, as shorter, if there is significantly more coarticulation than occurs with other consonants. If this is correct, then the vowel should be acoustically highly nasalized when the nasal is short, reflecting the true length of the nasal. Note that this would be consistent with the results for $C\tilde{V}C$ words in Beddor et al. (2013).

There is also evidence that may argue against the Expandability Hypothesis. It has been found that vowels preceding voiced fricatives are longer than vowels preceding voiced stops, while vowels preceding voiceless fricatives are somewhat longer than those preceding voiceless stops (Umeda, 1975; Peterson and Lehiste, 1960).[18] Furthermore, the voicing effect has been reported to be larger for fricatives than for stops (e.g., House and Fairbanks, 1953; House, 1961). Although our production experiment was not designed to explicitly test fricatives against stops, our results are in line with these findings. In our data, vowel durations were longest before voiced fricatives, and a larger voicing effect was found for fricatives than stops (91 ms $\triangle V$, versus 56 ms). However, the Expandability Hypothesis predicts that preceding vowel durations should be similar for voiced stops and fricatives, given that voiced fricatives were only 15 ms longer than voiced stops on average. It also predicts that vowels should be shorter before voiceless fricatives than voiceless stops, given that voiceless fricatives were about 29 ms longer than voiceless stops. A possible explanation could be that different consonants have different preferred C/V durations, thus exhibiting steeper or shallower curves for duration differences as a function of speaking rate.

[18] Umeda (1975) also finds that vowel duration preceding nasals is sometimes longer than before voiced stops, sometimes shorter, depending on the vowel. While vowels before voiceless fricatives tend to be intermediate in duration between voiceless stops and nasals, low vowels are actually longer before voiceless fricatives than before nasals. In a production experiment with Russian speakers, Kavitskaya (2002) finds that the *difference* in vowel duration between open and closed syllables is smallest before voiced fricatives, consistent with the other two studies. However, she also finds that voiceless stops have the next smallest difference, followed by voiceless fricatives, voiced stops, and nasals, with liquids showing the largest difference. The apparently variable behavior of nasals, voiceless stops and voiceless fricatives suggests that a number of interacting factors affect nucleus duration.

There is a suggestion of this in Figure 9, where there appear to be differences in the slopes of both ΔC and ΔV between the stop minimal pairs and the fricative minimal pairs. Allowing the C/V ratio to vary by phoneme class is undesirable from the perspective of theoretical economy. However, this hypothesis is testable, as are other aspects of the Expandability Hypothesis.

FIGURE 9 ABOUT HERE

The Expandability Hypothesis as developed here was designed for consistency with an already very large experimental literature, thus many of its predictions are actually postdictions. Nevertheless, we have offered a number of speculations that can, in principle, be tested. Among these are the hypothesis that onsets are largely excluded from rate/duration targets, that longer vowels in CVN words are highly nasalized, and that less nasalization in VNC sequences is correlated with longer VN durations. The competing constraints model also offers the hypothesis that significant differences in obstruent duration can occur without apparent compensation on vowels that are inherently short, or vowels that follow consonants at the beginning of the syllable (see Appendix B). Additional predictions about differences in effect size across final, medial, and initial position cannot be entirely determined by comparing across heterogeneous studies, but require carefully controlled experimentation to assess. More detailed information about gestural coordination between vowels and specific following consonants is also needed to fine-tune model predictions.

6 Summary & Conclusions

In much modern work, the voicing effect tends to be described in simplified terms, as a regular, quasi-universal, phonetically-driven phenomenon. In English, preceding vowel duration is often said to play a contrastive role for word-final stops (e.g., Klatt 1976). A small subset set of the literature – work that demonstrates either strong categorical perception (e.g., Raphael, 1972) or large vowel duration differences in production (e.g., Mack, 1982) – is most frequently cited. Such studies are primarily conducted using monosyllabic single-word stimuli in a laboratory setting, where speaking rate is much slower than for normally produced speech (production), and cues to stop identity are significantly impoverished, if not missing altogether (perception).

Yet it has been known for some time that vowel duration differences can be quite small in continuous speech, in polysyllabic words, across a syllable boundary, and phrase-medially (e.g., Umeda, 1975). Additionally, lax, unstressed, or otherwise inherently short vowels show little to no voicing effect even in laboratory speech (e.g., Peterson and Lehiste, 1960). In this shorter range of durations voiced and voiceless obstruents are of comparable length. Only at longer durations does the discrepancy between the two grow larger, as the voiced obstruent lengthens less and less. Voicing and obstruent duration are correlated, especially in the longer duration range, and analysis of the Buckeye Corpus showed that consonant duration was a better predictor of vowel duration.

In production studies that manipulate speaking rate it has been shown that voiceless obstruents, in both word-initial pre-stressed (VOT, e.g., Miller and Volaitis, 1989), and word-medial post-stress (closure duration, e.g., Port, 1976) position, are longer than voiced, with that difference increasing as speaking rate decreases. We extended that finding to coda position, demonstrating that the difference in vowel duration increased in step with the inverse duration difference for obstruents.[19] Using paired data, we were able to show that the magnitude of the "voicing" effect depended on obstruent duration across the board, while voicing was only significant at the slower rates (i.e., when it was significantly correlated with duration).

Aspiration, and actual voicing, have also been shown to be stronger cues to "voicing" than preceding vowel duration (Wardrip-Fruin, 1982; Repp and Williams, 1985). Furthermore, depending on the type of stimuli, vowel duration may have no effect on phoneme identification at all (Revoile et al., 1982). Obstruent duration itself has been shown to affect voicing perception in final position (Denes, 1955; Raphael, 1981; Repp and Williams, 1985), just as it does in

[19] In a similar study, Ko (2018) found that duration differences between voiced and voiceless obstruents, and between their preceding vowels, both increased with decreasing speaking rate. However, of the three speaking rates, the "normal" and "fast" conditions were largely the same, and duration differences were not analyzed as paired (voiced, voiceless) data.

word-medial position (Port and Dalby, 1982).

This body of results argues against preceding vowel duration as a primary cue to the voiced/voiceless contrast in English. Indeed, it strongly suggests that vowel duration directly affects the perception of obstruent duration, not voicing itself. We have offered a proposal that fits this larger range of findings. Namely, that the voicing effect in English is the result of the inherently low elasticity of voiced obstruents, and that segment durations, in general, are determined by the components of the Expandability Hypothesis, reproduced below.

(2) The Expandability Hypothesis

All segments have a characteristic elasticity that determines their resistance to lengthening

Resistance to lengthening increases with increasing duration for all segments

Lower elasticity equates with a more rapid increase in resistance

Relative resistance determines the distribution of duration across the syllable

The inverse correlation between obstruent duration and vowel duration, and its dependence on speaking rate, are attributed to a type of compensatory effect (see also Massaro and Cohen, 1983; Campbell, 1992), but not one based on syllable isochrony. Our competing constraints model of segment timing allows for "imperfect compensation", which appears to be the rule in language generally, rather than the exception (e.g., Browman and Goldstein, 1988; Krivokapić, 2020).

This model provides a proof of concept for deriving the voicing effect from a set of general-purpose timing constraints. The fact that the voicing effect is larger in fricatives than in stops cannot be explained under our account without allowing for a phoneme-class specific C/V parameter, however, we still cover much more empirical ground than explanations of the voicing effect that are based on actual vocal fold vibration, or articulatory effort. It is also worth noting that competing explanations (described in Section 1) have not attempted to explain this difference (while most don't even mention it). Whereas, we are able to unify the treatment of the contrast across word and syllable position, and draw connections between effects based on differences of consonant elasticity, and those based on differences of vowel elasticity. Our explanation for the voicing effect also has ramifications for theories of contrastive features.

6.1 Contrast and Allophony

Throughout this paper the relevant obstruent contrast in American English has been referred to as one of voicing. This is in spite of the fact that it is precisely because phonetic voicing is often absent from "voiced" stops that preceding vowel duration can be discussed as a possible cue to contrast. Clearly, the presence or absence of vocal fold vibration is not always necessary, or even sufficient, for phoneme identification. In order for the contrast to be described as one of voicing, it is necessary to treat the phonological voicing feature as distinct from the phonetic feature of the same name. The first is transformed to the second via a series of allophonic rules. For example, in absolute initial position the */-voice/* stop becomes *[+spread glottis]*, while the */+voice/* stop may become *[-voice].* In final position, a */-long/* vowel preceding a */+voice/* stop becomes *[+long]*.

However, we have seen that apparent vowel lengthening varies considerably as a function of speaking rate, sentence and word position, stress, and other factors (e.g., Crystal and House 1988; Umeda 1975). Most importantly, longer vowels correlate with shorter consonants, and voiced obstruents tend, cross-linguistically, to be shorter than their voiceless counterparts. The apparent physiological difficulty of maintaining the necessary conditions for voicing over extended closure periods has been proposed as an explanation for this tendency (e.g. Ohala, 1983, 2011). Nevertheless, it is possible, by virtue of greater articulatory effort, to maintain voicing if desirable, at least up to a point. Partial, or total, devoicing is also a possible outcome. Thus, whether or not voiced obstruents are actually shorter than voiceless obstruents is language specific. The fact that "voiced" stops in English are now frequently devoiced means that the observed duration differences are no longer the direct result of physiological constraints, but of what has become an underlyingly specified property of the segment. The fact that the difference in behavior between voiced and voiceless obstruents is only observable at long durations means that the specification is not for absolute duration, but for something that quantifies resistance to lengthening. The large voicing effect in English, we argue, is due to the voiced segment being pushed well beyond its preferred duration. Our claim is that vowel duration differences emerge directly from these elasticity differences. Therefore, we also conclude that vowel duration is *not* a feature that is specified is English, either at the phonological or phonetic level.

Although categorical perception effects have been demonstrated for the vowel duration cue, this is not particularly noteworthy, given that the number of acoustic cues to the contrast that listeners are able to exploit has been shown to be quite large. Duration and intensity of voicing, aspiration, and F0 contour, length of vowel formant transitions with respect to steady state duration (Fitch, 1981), F1 offset frequency (Crowther and Mann, 1992), speed of jaw lowering, and jaw offset position (Van Summers, 1987) all differ consistently between the two stop types in final position. In medial post-stress position, consistent differences have also been found in the timing of vocalic voice offset, and the signal decay time (Lisker, 1986), which should apply to final position as well. Furthermore, it is well known that cues can be "traded off" with one another. That is, while a long enough closure duration can cue a "voiceless" stop on its own, a shorter closure in tandem with a shortened vowel can also do so (e.g., Kohler, 1979, 1984; Fitch, 1981; Lisker, 1986; Van Summers, 1987; Bailey and Summerfield, 1980; Klatt, 1976; Malécot, 1968). Yet absolute vowel duration, and not closure duration or formant transition information, is frequently characterized as a phonological "voicing" feature, even though the latter two cues have been shown to influence perception to the same, or an even greater, degree. This may be due, in large part, to the privileging of 'prominent' contexts in phonological theory.

6.2 Prominence

While the phonetic realization of underlyingly contrastive features is assumed to vary by context, the most prominent environment, usually initial pre-stress position, is assumed to most faithfully reflect those features. Not only that, but features are said to be enhanced, or more strongly signaled, in such contexts (e.g, Kingston and Diehl, 1994). Conversely, observed enhancement is taken to indicate features that are "controlled", or underlyingly specified, as opposed to being supplied by context-sensitive rules (e.g., Ohala, 1981). Enhancement can be realized as an increase in acoustic amplitude, an increase in size of articulatory gestures, and/or an increase in gestural, and thus, segmental, duration (e.g., Beckman et al. 2013). In addition to making individual features more salient, enhancement is also assumed to be a mechanism for increasing discriminability between the members of a phonemic contrast (e.g., De Jong, 1995; Cho, 2016; Cho and Jun, 2000). For the above reasons, slower than normal speaking rate is considered to be an enhancement mechanism that should lead to lengthening, *but only of contrastively specified features* (e.g., Solé, 2007).

Underspecification theory applied to laryngeal contrasts typically makes use of the following privative features: [spread glottis], [voice], and [constricted glottis] (e.g., Kim, 1970; Iverson and Salmons, 1995). This system yields three possible two-way contrast systems, one for each of the features, with the second member always unspecified.The phonetically voiceless stops in French and Thai fail to lengthen significantly with decreased speaking rate, and are therefore taken to be unspecified for laryngeal features, while the phonetically short lag/voiced stops in English are the unspecified member of the contrast[20] (Kessinger and Blumstein, 1997; Beckman et al., 2013).

In the same vein, an observed interaction between a given phonetic cue, and any variable that affects duration, is taken to indicate that the cue is an inherent part of the contrast. It has been argued that vowel duration is purposefully manipulated by speakers to enhance the laryngeal contrast of the following obstruent, based on the following set of results: that the effect of stress is smaller for pre-voiceless than pre-voiced vowels in English (De Jong, 2004); that /-long/ pre-voiced vowels lengthen less than they would otherwise, in order to avoid overlapping with /+long/ pre-voiceless vowels, and preserve an existing long versus short vowel distinction in German (Braunschweiler, 1997); that vowel duration differences preceding voiced versus voiceless segments are greater for long vowels than for short vowels in English (Peterson and

[20] English is usually described as a [spread glottis]/Ø system, although this is not uncontroversial.

Lehiste, 1960); that the difference in duration between the stressed vowel in a monosyllabic word and the same vowel in a bisyllabic word is larger (by percentage) for syllables closed by voiced stops than those closed by voiceless stops (Van Summers, 1987; De Jong, 1991; Crowther and Mann, 1992; Raphael, 1975; Smith, 2002; Klatt, 1973); that the vowel shortening effect of affixation is greater (both absolutely, and proportionally) for a voiced-final stem than for a voiceless final (Lehiste, 1972).

In this paper, however, we have conceptualized stress, prosodic boundary marking, and speaking rate simply as external forces which, among others, can act to lengthen segments. Under our account, all segments are subject to such lengthening and shortening pressures. How much lengthening or shortening actually occurs for individual segments, however, is governed by the interactions of all such constraints, some of which are more highly weighted than others. The apparently asymmetric effects on voiced versus voiceless syllables do not need to be explained as the result of speaker effort to avoid phonetic ambiguity, or to maintain a specific range of phonetic values. They follow directly from these two premises: that the voicing effect derives from differences in segment elasticity; and that the resulting differences in duration increase with increasing duration. Characterizing the voicing effect as a consequence of on-line timing adjustments (to which multiple factors can contribute) is therefore more parsimonious, and more explanatorily adequate, than the hypothesis that there is both a grammatical rule of vowel lengthening, and a set of deliberate adjustments made to preserve the output of that rule. Note that this analysis requires elasticity to be underlyingly specified. This is not the same, however, as an underlying specification for an abstract voice feature. In the first place, all segments are assumed to have their own characteristic elasticity. Furthermore, a specification of this kind is necessary for independent reasons: to account for the differing degrees to which segments respond to changes in speaking rate. Finally, relative duration values within a word cannot be derived from a /voice/ feature, or even a long/short duration feature, as they depend on potentially complex interactions between all the segments within a word.

If ‘prominent’ contexts (such as slow speaking rate) do not actually enhance contrastive features, then the realization of the features in such contexts should not necessarily be taken as underlying. Doing so, in fact, requires potentially extensive transformations to arrive at the more frequent, non-prominent contexts of normal speech. If we reverse this relation, then very slow hyper-articulated speech is the exception, rather than the rule, and intense aspiration and especially long durations are derived from features that are more typical of the contrast in general. Large differences in preceding vowel duration are, almost exclusively, the product of atypical speech and therefore, in our view, should be considered the least central to the “voicing” contrast, not the most. This flipped view of contrast offers an intriguing avenue for future research.

7 References

Allen, J. Sean, and Joanne L. Miller. 1999. Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *The Journal of the Acoustical Society of America* 106:2031–2039.

Bailey, Peter J., and Quentin Summerfield. 1980. Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance* 6:536 – 563.

Beckman, Jill, Michael Jessen, and Catherine Ringen. 2013. Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics* 49:259–284.

Beddor, Patrice Speeter. 2009. A Coarticulatory Path to Sound Change. *Language* 85:785–821.

Beddor, Patrice Speeter, Kevin B. McGowan, Julie E. Boland, Andries W. Coetzee, and Anthony Brasher. 2013. The time course of perception of coarticulation. *The Journal of the Acoustical Society of America* 133:2350–2366.

Beguš, Gašper. 2017. Effects of ejective stops on preceding vowel duration. *The Journal of the Acoustical Society of America* 142:2168–2184.

Belasco, Simon. 1958. Variations in Vowel Duration: Phonemically or Phonetically Conditioned? *The Journal of the Acoustical Society of America* 30:1049–1050.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand, and Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60:92–111.

Berkovits, Rochele. 1993. Utterance-final lengthening and the duration of final-stop closures. *Journal of Phonetics* 21(4):479 – 489.

Boersma, Paul, and David Weenink. 2009. Praat: Doing phonetics by computer (Version 6.0.36). computer program. URL `http://www.fon.hum.uva.nl/praat/`.

Braunschweiler, Norbert. 1997. Integrated cues of voicing and vowel length in German: A production study. *Language and Speech* 40:353–376.

Browman, Catherine P., and Louis M. Goldstein. 1986. Towards an articulatory phonology. *Phonology* 3:219–252.

Browman, Catherine P., and Louis M. Goldstein. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45:140–155.

Browman, Catherine P., and Louis M. Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. In *Papers in laboratory phonology I: Between the grammar and the physics of speech*, ed. John Kingston and Mary E. Beckman, 341–376. Cambridge: Cambridge University Press.

Byrd, Dani, Abigail. R. Kaun, Shrikanth Narayanan, and Elliot Saltzman. 2000. Phrasal signatures in articulation. In *Papers in laboratory phonology v*, ed. M. B. Broe and J. B. Pierrehumbert, 70–87. Cambridge: Cambridge University Press.

Byrd, Dani, Sungbok Lee, Daylen Riggs, and Jason Adams. 2005. Interacting effects of syllable and phrase position on consonant articulation. *The Journal of the Acoustical Society of America* 118:3860–3873.

Byrd, Dani, and Elliot Saltzman. 1998. Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics* 26:173–199.

Byrd, Dani, and Elliot Saltzman. 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31:149–180.

Cambier-Langeveld, Gerda Martina. 1997. The domain of final lengthening in the production of Dutch. *Linguistics in the Netherlands* 14:13–24.

Cambier-Langeveld, Gerda Martina. 2000. *Temporal marking of accents and boundaries*. Den Haag: Holland Academic Graphics.

Campbell, W. Nick. 1992. Syllable-based segmental duration. *Talking machines: Theories, models, and designs* 211–224.

Catford, John Cunnison. 1977. *Fundamental problems in phonetics*. Midland Books.

Chen, Matthew. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22:129–159.

Cho, Taehong. 2016. Prosodic boundary strengthening in the phonetics–prosody interface. *Language and Linguistics Compass* 10:120–141.

Cho, Taehong, and Sun-Ah Jun. 2000. Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean. *UCLA Working Papers in Phonetics* 57–70.

Cho, Taehong, and Patricia Keating. 2009. Effects of initial position versus prominence in English. *Journal of Phonetics* 37:466–485.

Coretta, Stefano. 2019. An exploratory study of voicing-related differences in vowel duration as compensatory temporal adjustment in Italian and Polish. *Glossa: a journal of general linguistics* 4.

Crowther, Court S., and Virginia Mann. 1992. Native language factors affecting use of vocalic cues to final consonant voicing in English. *The Journal of the Acoustical Society of America* 92:711–722.

Crystal, Thomas H, and Arthur S House. 1982. Segmental durations in connected speech signals: Preliminary results. *The Journal of the Acoustical Society of America* 72:705–716.

Crystal, Thomas H, and Arthur S House. 1988. Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America* 83:1553–1573.

Davis, Stuart, and W. Van Summers. 1989. Vowel length and closure duration in word-medial VC sequences. *Journal of Phonetics* 17:339–353.

De Jong, Kenneth J. 1991. An articulatory study of consonant-induced vowel duration changes in English. *Phonetica* 48:1–17.

De Jong, Kenneth J. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America* 97:491–504.

De Jong, Kenneth J. 2004. Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics* 32:493–516.

Delattre, Pierre. 1962. Some factors of vowel duration and their cross-linguistic validity. *The Journal of the Acoustical Society of America* 34:1141–1143.

Delattre, Pierre. 1966. A comparison of syllable length conditioning among languages. *IRAL-International Review of Applied Linguistics in Language Teaching* 4:183–198.

Denes, Peter. 1955. Effect of duration on the perception of voicing. *The Journal of the Acoustical Society of America* 27:761–764.

Derr, Marcia A, and Dominic W. Massaro. 1980. The contribution of vowel duration, F0 contour, and frication duration as cues to the /juz/-/jus/ distinction. *Perception & Psychophysics* 27:51–59.

Elert, Claes Christian. 1965. *Phonologic studies of quantity in Swedish: Based on material from Stockholm speakers*. Almqvist och Wiksell.

Fidelholtz, James L. 1975. Word frequency and vowel reduction in English. In *Papers from the Eleventh Regional Meeting of the Chicago Linguistic Society*, volume 11, 200–213.

Fischer, Rebecca M., and Ralph N. Ohde. 1990. Spectral and duration properties of front vowels as cues to final stop-consonant voicing. *The Journal of the Acoustical Society of America* 88:1250–1259.

Fitch, Hollis Leslie. 1981. Distinguishing temporal information for speaking rate from temporal information for intervocalic stop consonant voicing. Technical report, Haskins Laboratory.

Flege, James. 1979. Phonetic interference in second language acquisition. Doctoral Dissertation, Indiana University.

Fosler-Lussier, Eric, and Nelson Morgan. 1999. Effects of speaking rate and word frequency on pronunciations in conversational speech. *Speech Communication* 29:137–158.

Fougeron, C., and P. A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America* 101:3728–3740.

Fox, Robert A., and Dale Terbeek. 1977. Dental flaps, vowel duration and rule ordering in American English. *Journal of Phonetics* 5:27–34.

Fulop, Sean A, and Hannah J Scott. 2021. Vowel duration in the Buckeye corpus. *The Journal of the Acoustical Society of America* 150:A69–A69.

Gahl, Susanne. 2009. Homophone duration in spontaneous speech: A mixed-effects model. Technical Report 5.

Gahl, Susanne, and Harald Baayen. 2022. Time and thyme again: Connecting spoken word duration to models of the mental lexicon .

Gahl, Susanne, Yao Yao, and Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66:789–806.

Gareth, James, Witten Daniela, Hastie Trevor, and Tibshirani Robert. 2013. *An introduction to statistical learning: with applications in r*. Spinger.

Halle, Morris, and Kenneth Stevens. 1967. Mechanism of glottal vibration for vowels and consonants. *The Journal of the Acoustical Society of America* 41:1613–1613.

Hardcastle, William J. 1985. Some phonetic and syntactic constraints on lingual coarticulation in stop consonant sequences. *Speech Communication* 4:247–263.

Harris, MS, and Noriko Umeda. 1974. Effect of speaking mode on temporal factors in speech: Vowel duration. *The Journal of the Acoustical Society of America* 56:1016–1018.

Hillenbrand, James M., Michael J. Clark, and Robert A. Houde. 2000. Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America* 108:3013–3022.

Hillenbrand, James M., Dennis R. Ingrisano, Bruce L. Smith, and James E. Flege. 1984. Perception of the voiced–voiceless contrast in syllable-final stops. *The Journal of the Acoustical Society of America* 76:18–26.

Hofhuis, E., Carlos Gussenhoven, and Toni Rietveld. 1995. Final lengthening at prosodic boundaries in Dutch. volume 1, 154–157. Stockholm: Stockholm University.

Hogan, John T., and Anton J. Rozsypal. 1980. Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *The Journal of the Acoustical Society of America* 67:1764–1771.

Hooper, Joan B. 1976. Word frequency in lexical diffusion and the source of morphophonological change. In *Current progress in historical linguistics*, ed. William Christie, 96–105. Amsterdam: North Holland.

House, Arthur S. 1961. On vowel duration in English. *The Journal of the Acoustical Society of America* 33:1174–1178.

House, Arthur S, and Grant Fairbanks. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America* 25:105–113.

Hyman, Larry. 2019. *A theory of phonological weight*. De Gruyter Mouton.

Iverson, Gregory K., and Joseph C. Salmons. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12:369–396.

Javkin, Hector. 1977. Phonetic universals and phonological change. Doctoral Dissertation, U.C. Berkeley.

Jessen, Michael. 2001. *Distinctive feature theory*, volume 2, chapter Phonetic implementation of the distinctive auditory features [voice] and [tense] in stop consonants, 237–294. Mouton de Gruyter Berlin.

Johnson, Keith. 2004. Massive reduction in conversational American English. In *Proceedings of the Workshop on Spontaneous Speech: Data and Analysis*, 29–54.

Jurafsky, D., A. Bell, M. Gregory, and W. D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In *Frequency and the emergence of linguistic structure*, ed. Joan L. Bybee and Paul J. Hopper, number 45 in Typological studies in language, 229–254. Amsterdam: John Benjamins.

Jurafsky, Daniel, Alan Bell, Eric Fosler-Lussier, Cynthia Girand, and William Raymond. 1998. Reduction of English function words in Switchboard. In *Fifth International Conference on Spoken Language Processing*.

Katsika, Argyro. 2016. The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of phonetics* 55:149–181.

Kavitskaya, D. 2002. *Compensatory lengthening: Phonetics, phonology, diachrony*. London: Routledge.

Keating, Patricia A. 1979. A phonetic study of a voicing contrast in Polish. Doctoral Dissertation, Brown University.

Keating, Patricia A. 1985. Universal phonetics and the organization of grammars. In *Phonetic* *linguistics: Essays in honor of Peter Ladefoged*, ed. Victoria A. Fromkin, 115–132. Orlando: Academic Press.

Kessinger, Rachel H., and Sheila E. Blumstein. 1997. Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics* 25:143–168.

Kim, Chin-Wu. 1970. A theory of aspiration. *Phonetica* 21:107–116.

Kim, Heejin, and Jennifer Cole. 2005. The stress foot as a unit of planned timing: evidence from shortening in the prosodic phrase. In *Ninth European Conference on Speech Communication and Technology*.

Kim, Sahyang, and Taehong Cho. 2012. Prosodic strengthening in the articulation of English /æ/. *Studies in Phonetics, Phonology and Morphology* 18:321–337.

Kingston, J., and R. L. Diehl. 1994. Phonetic Knowledge. *Language* 70:419–454.

Klatt, Dennis H. 1973. Interaction between two factors that influence vowel duration. *The Journal of the Acoustical Society of America* 54:1102–1104.

Klatt, Dennis H. 1975. Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics* 3:129–140.

Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America* 59:1208–1221.

Kluender, Keith R., Randy L. Diehl, and Beverly A. Wright. 1988. Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics* 16:153–169.

Ko, Eon-Suk. 2018. Asymmetric effects of speaking rate on the vowel/consonant ratio conditioned by coda voicing in English. *Phonetics and Speech Sciences* 10:45–50.

Kohler, K. J. 1979. Dimensions in the perception of fortis and lenis consonants. *Phonetica* 36:332–343.

Kohler, Klaus J. 1984. Phonetic explanation in phonology: the feature fortis/lenis. *Phonetica* 41:150–174.

Kozhevnikov, Valeriĭ Aleksandrovich, and Liudmila Andreevna Chistovich. 1965. *Speech: Articulation and perception*. Nauka.

Krause, Sue Ellen. 1982. Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults. *The Journal of the Acoustical Society of America* 71:990–995.

Kristoffersen, Gjert. 2000. *The phonology of Norwegian*. Oxford: Oxford University Press on Demand.

Krivokapić, Jelena. 2020. *Prosodic theory and practice*, chapter Prosody in Articulatory Phonology. MIT Press. In press Cambridge, MA.

Kulikov, Vladimir. 2012. Voicing and voice assimilation in Russian stops. Doctoral Dissertation, University of Iowa.

Laeufer, Christiane. 1992. Patterns of voicing-conditioned vowel duration in French and English. *Journal of Phonetics* 20:411–440.

Lehiste, Ilse. 1972. The timing of utterances and linguistic boundaries. *The Journal of the Acoustical Society of America* 51:2018–2024.

Lisker, Leigh. 1957. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language* 33:42–49.

Lisker, Leigh. 1974. On "explaining" vowel duration variation. *Glossa* 8:233–246.

Lisker, Leigh. 1978. Rapid vs. Rabid: A catalogue of acoustic features that may cue the distinction. *Haskins Laboratories Status Report on Speech Research* 54:127–132.

Lisker, Leigh. 1986. "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29:3–11.

Luce, Paul A. 1986. Neighborhoods of Words in the Mental Lexicon. Research on Speech Perception. Technical Report No. 6. Technical report, Indiana University.

Luce, Paul A, and Jan Charles-Luce. 1985. Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *The Journal of the Acoustical Society of America* 78:1949–1957.

Mack, Molly. 1982. Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *The Journal of the Acoustical Society of America* 71:173–178.

Malécot, André. 1968. The force of articulation of American stops and fricatives as a function of position. *Phonetica* 18:95–102.

Massaro, Dominic W., and Michael M. Cohen. 1983. Consonant/vowel ratio: An improbable cue in speech. *Attention, Perception, & Psychophysics* 33:501–505.

Miller, Joanne L. 1981. *Perspectives on the study of speech*, chapter Effects of speaking rate on segmental distinctions, 39–74. Routledge.

Miller, Joanne L., and Thomas Baer. 1983. Some effects of speaking rate on the production of /b/ and /w/. *The Journal of the Acoustical Society of America* 73:1751–1755.

Miller, Joanne L., Kerry P. Green, and Adam Reeves. 1986. Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica* 43:106–115.

Miller, Joanne L, and Lydia E Volaitis. 1989. Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics* 46:505–512.

Moreton, Elliott. 2004. Realization of the English postvocalic [voice] contrast in F1 and F2. *Journal of Phonetics* 32:1–33.

Munhall, Kevin, Carol Fowler, Sarah Hawkins, and Elliot Saltzman. 1992. "Compensatory shortening" in monosyllables of spoken English. *Journal of Phonetics* 20:225–239.

Nagao, Kyoko, and Kenneth J. de Jong. 2007. Perceptual rate normalization in naturally produced rate-varied speech. *The Journal of the Acoustical Society of America* 121:2882–2898.

Nam, Hosung, and Elliot Saltzman. 2003. A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th international congress of phonetic sciences*, volume 1, 2253–2256.

Nittrouer, Susan. 2004. The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *The Journal of the Acoustical Society of America* 115:1777–1790.

O'Dell, Michael, and Tommi Nieminen. 1999. Coupled oscillator model of speech rhythm. In *Proceedings of the XIVth international congress of phonetic sciences*, volume 2, 1075–1078.

Ohala, J. J. 2011. Accommodation to the Aerodynamic Voicing Constraint and its Phonological Relevance. In *Proceedings of the 15th International Conference of Phonetic Sciences*, 64–67.

Ohala, John J. 1981. The Listener as a source of sound change. In *Parasession on language and behavior*, ed. C. S. Masek, R. A. Hendrick, and M. F. Miller, 178–203. Chicago: Chicago Linguistics Society.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In *The production of speech*, ed. P. F. MacNeilage, 189–216. New York: Springer.

O'Kane, Donal. 1978. Manner of vowel termination as a perceptual cue to the voicing status of postvocalic stop consonants. *Journal of Phonetics* 6:311–18.

Oller, D. Kimbrough. 1973. The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America* 54:1235–1247.

Peterson, Gordon E., and Ilse Lehiste. 1960. Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America* 32:693–703.

Pike, Kenneth L. 1945. *The intonation of american english*. University of Michigan Academic Press.

Pind, Jörgen. 1995. Speaking rate, voice-onset time, and quantity: The search for higher-order invariants for two Icelandic speech cues. *Perception & Psychophysics* 57:291–304.

Pitt, Mark A, Laura Dilley, Keith Johnson, Scott Kiesling, William Raymond, E Hume, and E Fosler-Lussier. 1997. Buckeye Corpus of Conversational Speech (2nd release). Department of Psychology, Ohio State University (Distributor), Columbus, OH, USA.

Pitt, Mark A, Keith Johnson, Elizabeth Hume, Scott Kiesling, and William Raymond. 2005. The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication* 45:89–95.

Pluymaekers, Mark, Mirjam Ernestus, and R. Harald Baayen. 2005. Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America* 118:2561–2569.

Port, Robert F. 1976. The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words. Doctoral Dissertation.

Port, Robert F. 1979. The influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics* 7(1):45–56.

Port, Robert F. 1981. Linguistic timing factors in combination. *The Journal of the Acoustical Society of America* 69:262–274.

Port, Robert F., and Fred Cummins. 1992. The English voicing contrast as velocity perturbation. In *Proceedings of the Second International Conference on Spoken Language Processing*, 1311–1314. Banff, Alberta, Canada.

Port, Robert F., and Jonathan Dalby. 1982. Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics* 32:141–152.

Priva, Uriel Cohen. 2010. Constructing typing-time corpora: A new way to answer old questions. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 32, 43–48.

Priva, Uriel Cohen. 2017. Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition* 160:27–34.

Raphael, Lawrence J. 1972. Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *The Journal of the Acoustical Society of America* 51:1296–1303.

Raphael, Lawrence J. 1975. The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English. *Journal of Phonetics* 3:25–33.

Raphael, Lawrence J. 1981. Durations and contexts as cues to word-final cognate opposition in English. *Phonetica* 38:126–147.

Raphael, Lawrence J., Michael F. Dorman, Frances Freeman, and Charles Tobin. 1975. Vowel and nasal duration as cues to voicing in word-final stop consonants: Spectrographic and perceptual studies. *Journal of Speech and Hearing Research* 18:389–400.

Raymond, William D, Mark Pitt, Keith Johnson, Elizabeth Hume, Matthew Makashay, Robin Dautricourt, and Craig Hilts. 2002. An analysis of transcription consistency in spontaneous speech from the Buckeye corpus. In *Seventh International Conference on Spoken Language Processing*.

Repp, Bruno H., and David R. Williams. 1985. Influence of following context on perception of the voiced–voiceless distinction in syllable-final stop consonants. *The Journal of the Acoustical Society of America* 78:445–457.

Revoile, S., J.M. Pickett, Lisa D. Holden, and David Talkin. 1982. Acoustic cues to final stop voicing for impaired- and normal- hearing listeners. *The Journal of the Acoustical Society of America* 72:1145–1154.

Rositzke, Harry A. 1939. Vowel-length in General American speech. *Language* 15:99–109.

Saltzman, Elliot, Hosung Nam, Jelena Krivokapic, and Louis Goldstein. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Proceedings of the 4th International Conference on Speech Prosody (Speech Prosody 2008), Campinas, Brazil*, 175–184.

Sanker, Chelsea. 2019. Influence of coda stop features on perceived vowel duration. *Journal of Phonetics* 75:43–56.

Sanker, Chelsea. 2020. A perceptual pathway for voicing-conditioned vowel duration. *Laboratory Phonology* 11.

Schwartz, Geoffrey. 2010. Phonology in the speech signal-Unifying cue and prosodic licensing. *Poznań Studies in Contemporary Linguistics* 46:499–518.

Selkirk, Elizabeth. 1982. *The structure of phonological representations part 2*, chapter The syllable, 337–383. Dordrecht: Foris.

Sharf, Donald J. 1962. Duration of post-stress intervocalic stops and preceding vowels. *Language and speech* 5:26–30.

Smith, Bruce L. 2002. Effects of speaking rate on temporal patterns of English. *Phonetica* 59:232–244.

Solé, Maria-Josep. 2007. *Experimental approaches to phonology*, chapter Controlled and mechanical properties in speech, 302–321. Oxford University Press.

Stetson, Raymond H. 1928. *Motor phonetics: A study of speech movements in action*. Dordrecht: Springer.

Stevens, K. N., and A. S. House. 1963. Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research* 6:111–128.

Summerfield, Q. 1981. Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 7:1074–1095.

Sweet, Henry. 1880. *A handbook of phonetics*. Clarendon Press Series. London: MacMillan and Co.

Tanner, James, Morgan Sonderegger, Jane Stuart-Smith, and SPADE Data Consortium. 2019. Vowel duration and the voicing effect across English dialects. *Toronto Working Papers in Linguistics* 41.

Toscano, Joseph C., and Bob McMurray. 2015. The time-course of speaking rate compensation: Effects of sentential rate and vowel length on voicing judgments. *Language, Cognition and Neuroscience* 30:529–543.

Turk, Alice E., and Stefanie Shattuck-Hufnagel. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics* 35:445–472.

Umeda, Noriko. 1975. Vowel duration in American English. *The Journal of the Acoustical Society of America* 58:434–445.

Van Heuven, Walter J.B., Pawel Mandera, Emmanuel Keuleers, and Marc Brysbaert. 2014. SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology* 67:1176–1190.

Van Summers, W. 1987. Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *The Journal of the Acoustical Society of America* 82:847–863.

Vatikiotis-Bateson, Eric. 1984. The temporal effects of homorganic medial nasal clusters. *Research in Phonetics* 4:197–233.

Viswanathan, Navin, Annie J. Olmstead, and M. Pilar Aivar. 2019. The use of vowel length in making voicing judgments by native listeners of English and Spanish: Implications for rate normalization. *Language and Speech* 63:436–452.

Volaitis, L. E., and J. L. Miller. 1992. Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America* 92:723–735.

Walsh, Thomas, and Frank Parker. 1981. Vowel length and voicing in a following consonant. *Journal of Phonetics* 9:305–308.

Wardrip-Fruin, Carolyn. 1982. On the status of temporal cues to phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants. *The Journal of the Acoustical Society of America* 71:187–195.

Wightman, C. W., S. Shattuck-Hufnagel, M. Ostendorf, and P. J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91:1707–1717.

Table 1

Utterance-Final: Simple Model: No interactions

	Estimate	Std. Error	df	t value	Pr(>|t|)
(intercept)	4.81	0.03	110	140	< 2e-16
Vowel Height	0.16	0.02	223	9.11	< 2e-16
Vowel Type	0.11	0.02	214	6.59	3.3e-10
Speaking Rate	0.67	0.02	70.8	29.6	< 2e-16
Voicing	0.04	0.02	210	2.30	0.02
Obstruent Type	0.02	0.02	160	1.33	0.18
Frequency	0.09	0.03	129	-3.01	0.003

Table 2
Utterance-Final: Simple Model with consonant duration.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(intercept)	4.80	0.04	85.6	127	< 2e-16
Vowel Height	0.16	0.02	192	10.7	< 2e-16
Vowel Type	0.12	0.01	186	8.74	1.4e-15
Speaking Rate	1.03	0.02	162	42.7	< 2e-16
Consonant Duration	-0.50	0.02	3152	-29.2	< 2e-16
Obstruent Type	-0.05	0.02	167	-3.51	5.7e-4
Frequency	-0.07	0.03	126	-2.59	.01

Table 3

Utterance-Final only: Consonant Duration model.

	Estimate	Std. Error	df	t value	Pr(>ltl)
(intercept)	-.02	0.03	114	-0.75	0.45
Voicing	-0.08	0.02	120	-3.59	4.8e-4
Speaking Rate	0.73	0.01	3264	54.13e-14	< 2e-16
Obstruent Type	-0.16	0.02	239	-10.4	< 2e-16
Frequency	0.02	0.03	217	0.72	0.47
Voicing: Rate	-0.001	0.01	3254	-0.08	0.94
Voicing: Frequency	0.01	256	217	0.55	0.58
Voicing: Obs. Type	0.02	0.02	239	1.03	0.30

Table 4

Mixed-Effects Linear Regression Model of vowel duration as a function of speaking rate and consonant duration and their interaction.

	Estimate	Std. Error	estimated df	t-value	p-value
(Intercept)	355.0	25.05	27.31	14.17	4.13e-14
rate.L	324.5	15.04	969.9	21.58	< 2e-16
rate.Q	42.49	15.16	968.7	2.802	0.005
Consonant Duration	-0.169	0.059	977.7	-2.879	0.004
rate.L:C Duration	-0.637	0.113	969.3	-6.657	2.03e-8

Table 5

Mixed-Effects Linear Regression Model of consonant duration as a function of speaking rate, voicing, and manner, with full interactions.

	Estimate	Std. Error	estimated df	t-value	p-value
(Intercept)	160.1	7.253	7.175	22.07	7.37e-8
rate.L	87.84	6.835	960.2	12.85	< 2e-16
voicing	-54.96	8.804	3.987	-6.243	0.003
manner	28.83	8.812	4.001	3.272	0.031
rate.L:voicing	-46.76	9.702	960.2	-4.819	1.67e-6
rate.L:manner	19.59	9.715	960.2	2.016	0.044

Table 6

Mixed-Effects Linear Regression Model of vowel duration difference as a function of consonant duration difference, manner and rate, with full interactions.

	Estimate	Std. Error	estimated *df*	*t*-value	*p*-value
(Intercept)	-57.37	12.34	6.891	-4.649	0.002
$\triangle C$	-0.217	0.064	475.8	-3.390	0.001
rate.Q	30.11	10.94	451.8	2.751	0.006
rate.L:$\triangle C$	-0.376	0.139	460.4	-2.701	0.007
rate.Q:$\triangle C$	-0.277	0.133	455.8	-2.081	0.038

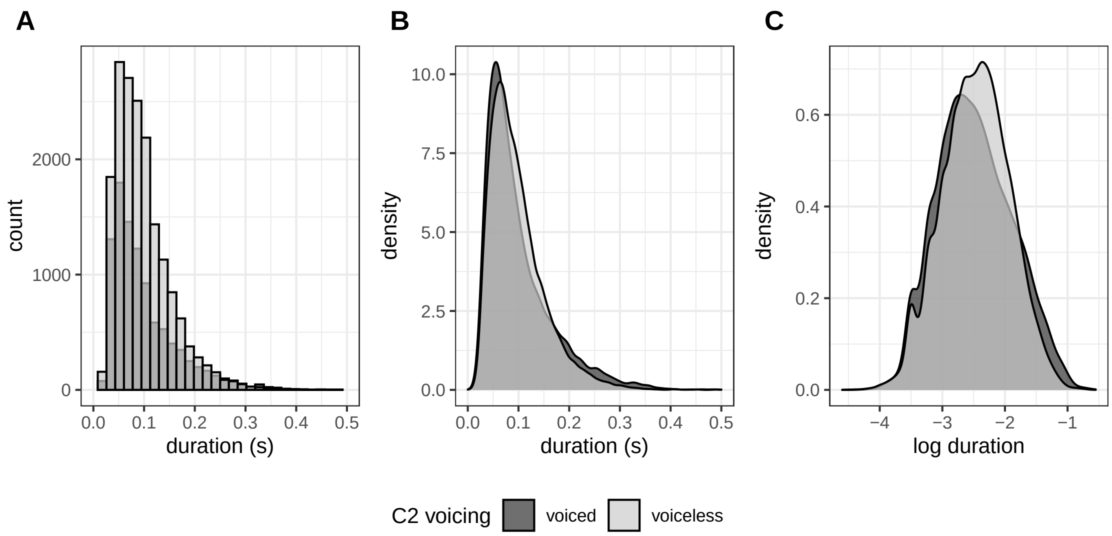


Figure 1
All CVC vowel durations

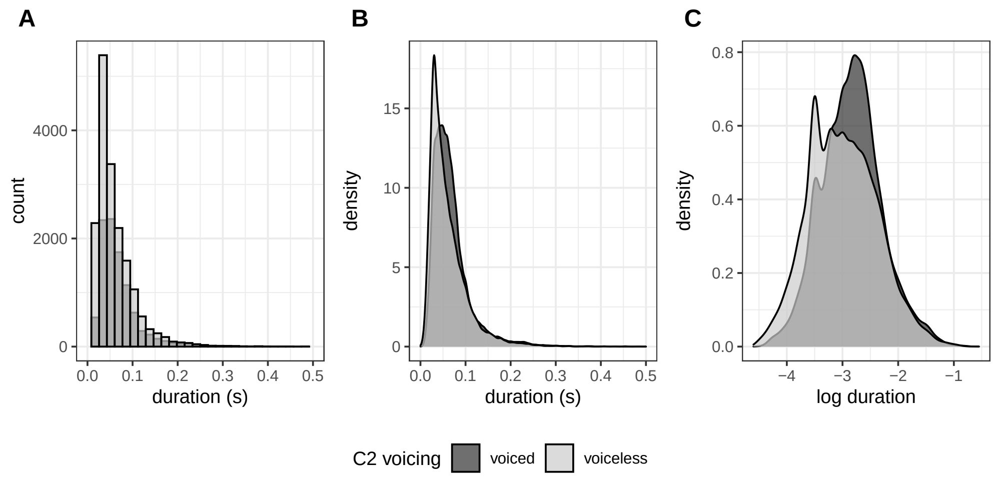


Figure 2
Obstruent durations (flaps excluded).

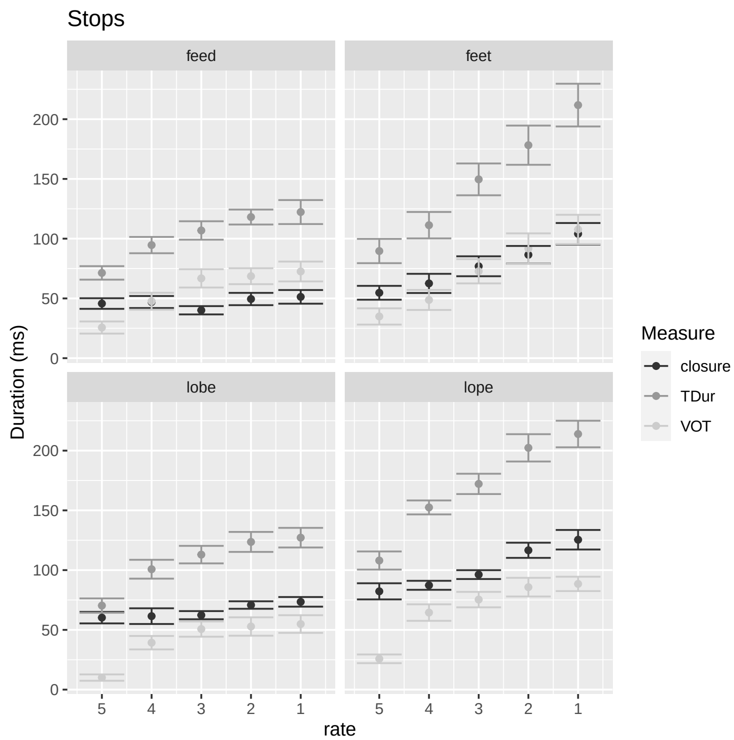


Figure 3

Closure duration, VOT and Total duration (TDur) for final stops as a function of repetition rate (decreasing from left to right). Means and standard error bars.

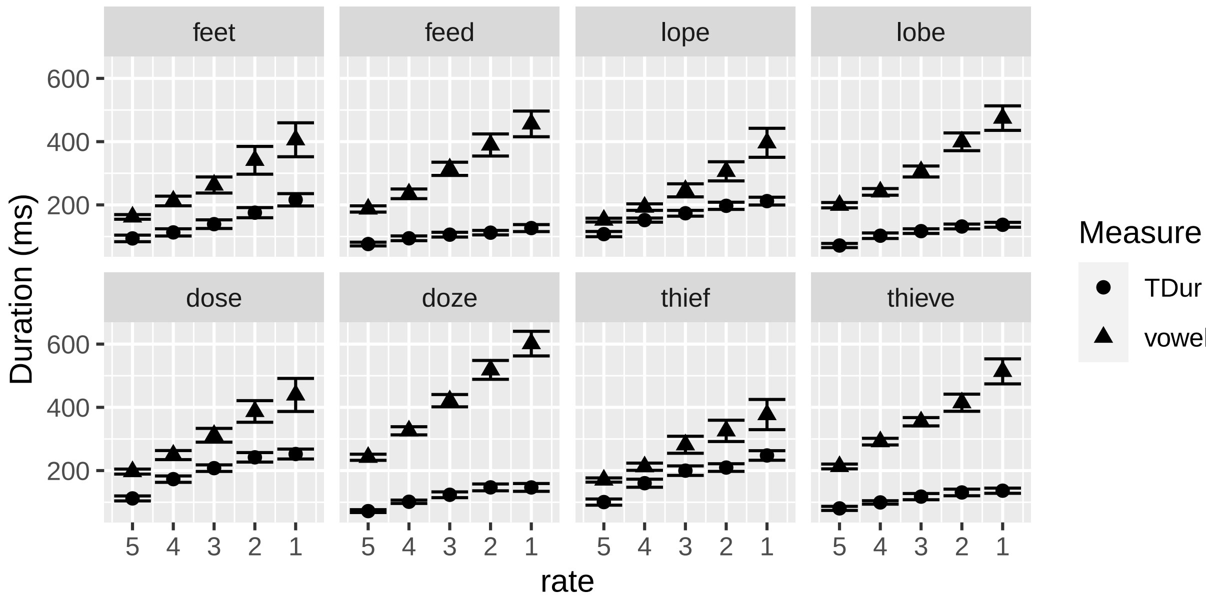


Figure 4
Total consonant duration, preceding vowel duration, and rhyme duration (V+C), as a function of repetition rate. Means and standard error bars.

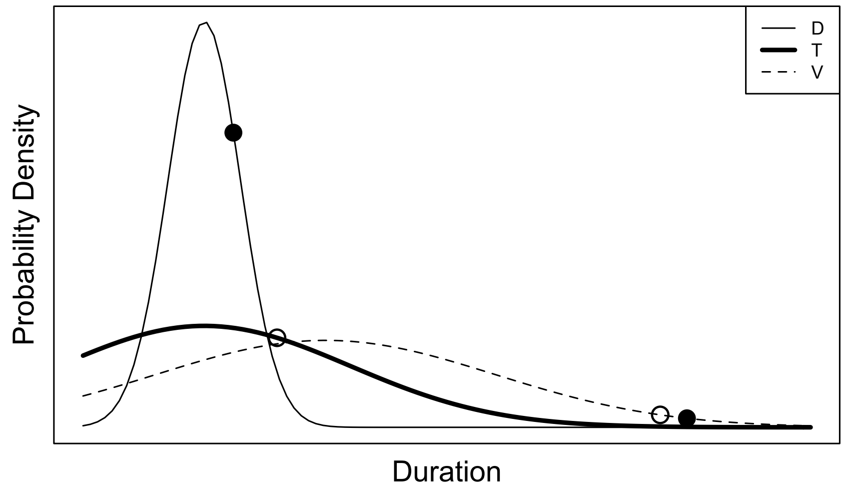


Figure 5
Probability densities for: voiced obstruent (solid); voiceless obstruent (thick solid); vowel (dashed). Open circles: VT syllable with target syllable duration of 330 ms. Filled circles: VD syllable with target syllable duration of 330 ms.

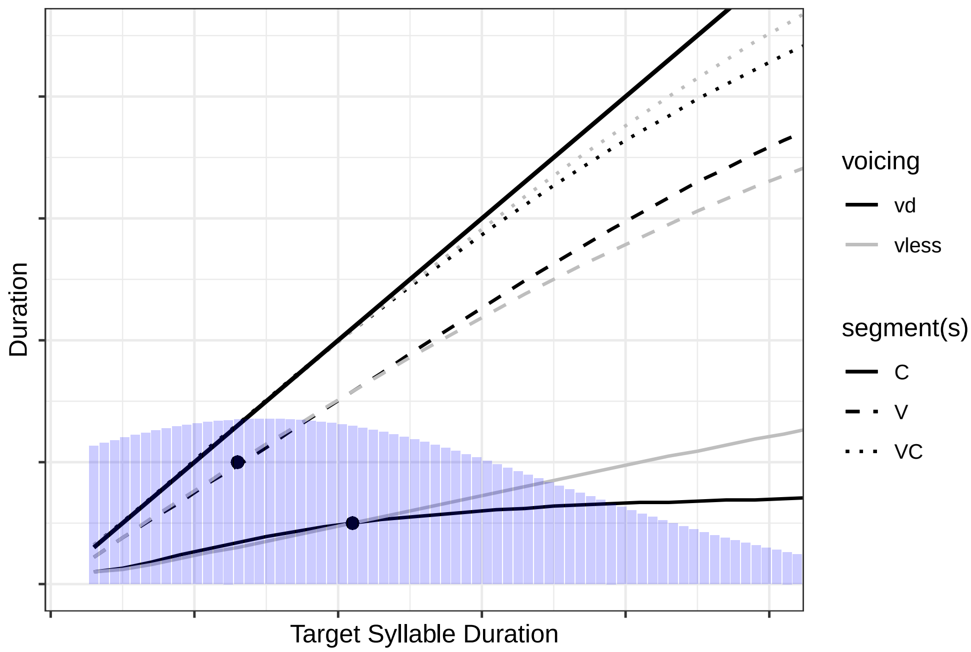
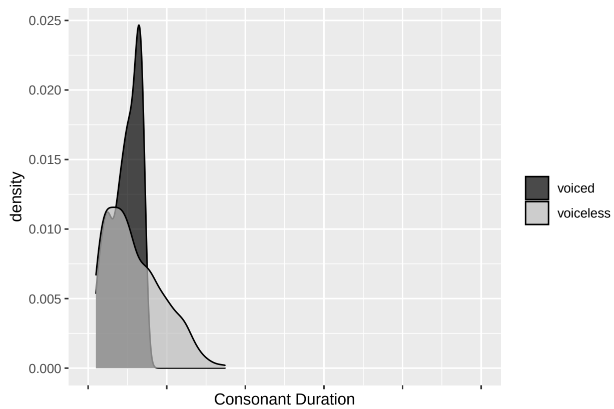


Figure 6

Behavior of the Competing Constraints Model of segment duration as a function of target syllable duration. Actual and target syllable duration are equal along the upper solid black line. Vertical blue bars show the distribution used to represent the durations/rates in the Buckeye Corpus (used for the corpus simulation).

(a) *Consonant Duration*



(b) *Vowel Duration*

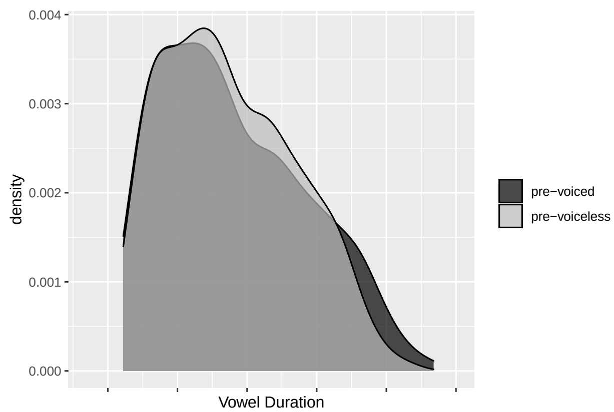


Figure 7

Corpus Simulation. Target syllable durations randomly sampled from the Normal distribution shown in Fig. 6.

(a) *Short vowel token* (b) *Long vowel token*

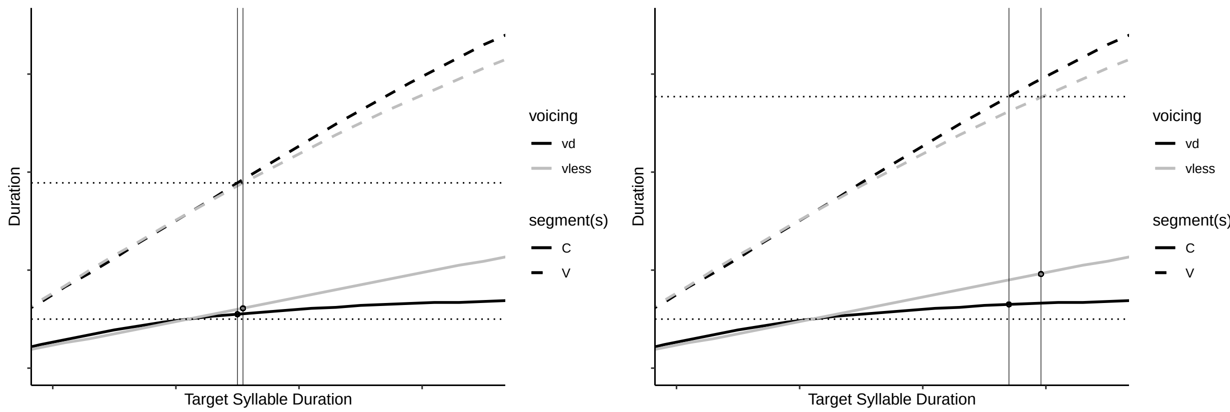


Figure 8

Competition Simulation: Observed vowel duration is marked by the upper dotted line in both figures. Vertical solid lines intersect expected target syllable duration (speaking rate), and expected stop duration. Left line: voiced stop coda; Right line: voiceless stop coda. The lower dotted line indicates the actual duration of the following stop.

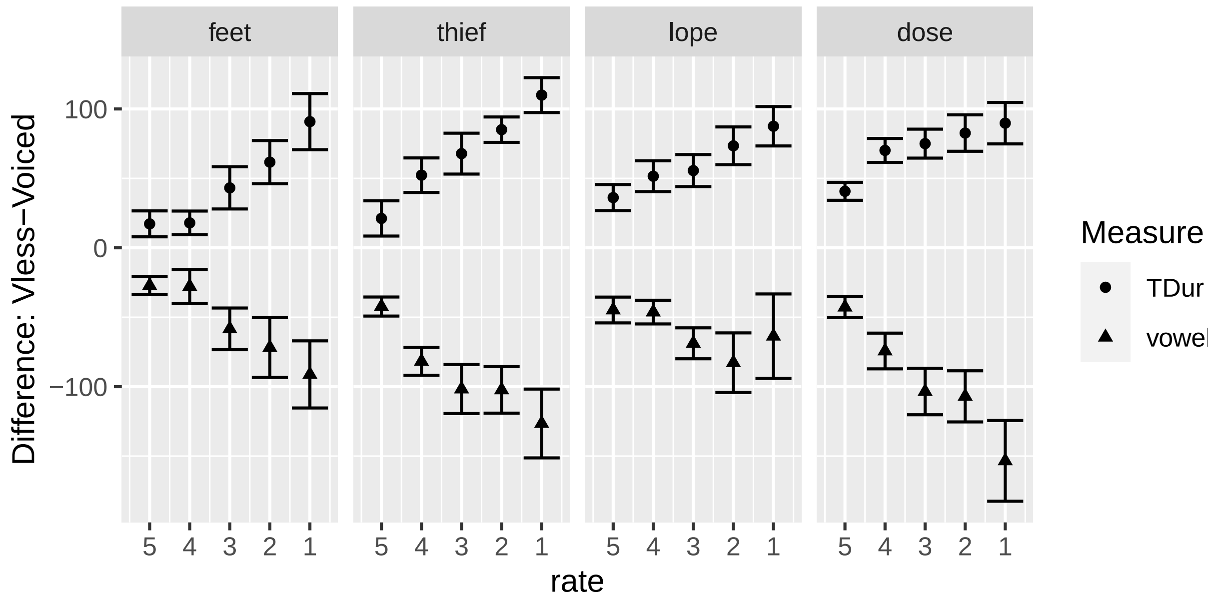


Figure 9

Production differences by voicing for each minimal pair.

Appendix A

Word Lists

CV Stop.

Voiced (3071 tokens; 83 unique words) , with individual counts: bad(111), bag(3), bed(14), big(154), bob(2), cab(3), cad(1), cod(1), code(2), could(206), cub(1), dab(2), dad(60), dead(11), did(232), died(8), dig(1), dog(18), dude(2), fed(4), feed(3), fog(4), food(20), gig(1), god(35), good(245), guide(1), had(384), head(16), hid(1), hide(4), hood(1), hub(1), hug(2), hyde(2), jed(1), jedd(2), job(104), kid(92), knob(1), lab(1), lag(1), laid(5), lead(8), league(8), led(2), leg(3), lied(1), load(2), loud(10), mad(5), made(63), med(4), meg(1), mid(4), mud(1), need(150), paid(31), pig(2), read(43), red(7), rid(1), ride(9), road(22), rob(2), rub(1), sad(5), said(311), shed(2), should(140), showed(5), side(32), sued(2), tag(2), ted(1), tied(1), todd(1), tub(1), tube(1), web(6), weed(1), wide(1), would(416)

Voiceless (8663 tokens; 173 unique words), with individual counts: back(341), beat(9), beep(1), bet(9), bike(1), bit(27), bite(1), boat(3), book(23), bought(9), buck(3), but(880), butt(2), cake(4), cap(2), cape(1), cat(2), caught(4), chalk(1), cheap(11), check(15), chick(1), chip(1), coke(1), cook(11), cop(9), cope(1), cup(1), cut(12), date(5), deck(4), deep(9), dip(1), dot(6), doubt(1), duck(1), duke(1), fake(5), fat(3), feet(4), fight(5), fit(10), folk(3), foot(1), fuck(1), gap(1), gate(2), get(315), got(156), gut(1), hate(14), heat(1), heck(14), height(1), hick(1), hip(2), hit(11), hook(4), hop(4), hope(29), hot(6), hype(2), jack(1), jeep(2), jet(1), jock(2), joke(6), keep(88), kick(3), knit(2), lack(12), lake(7), lap(1), late(6), let(28), light(5), like(2537), lock(10), look(149), lot(75), luck(1), luke(1), mac(1), make(225), map(2), meet(9), met(17), might(44), mike(3), mock(2), nap(1), neat(8), neck(4), net(1), night(14), nope(6), nose(5), not(322), note(1), nut(1), pack(5), peek(1), pet(1), pete(1), pick(31), pipe(4), poke(1), pop(1), pope(2), pot(1), psych(2), puck(1), put(2), rat(3), rate(2), rec(2), right(197), rock(6), rope(1), route(3), sake(3), sat(6), seat(1), set(13), shake(1), shape(7), sheet(2), ship(4), shit(4), shock(4), shoot(17), shop(6), shot(4), shut(1), sick(7), sit(39), site(3), soap(7), soup(5), suit(3), take(255), talk(125), tap(1), tape(11), taught(8), tech(5), that(1413), thick(1), this(49), thought(91), tight(1), tip(4), took(88), top(25), type(48), vote(20), wait(11), wake(2), week(78), weight(1), wet(2), whack(3), what(346), whip(2), white(6), wick(1), woke(2), wreck(1), wright(1), write(13), wrote(6), yet(22), zip(4)

7.0.1 CV Fricative

Voiced (4912 tokens; 69 unique words), with individual counts: b's(4), boys(32), c's(4), cahs(1), cause(53), cave(1), cheese(4), choose(15), chose(3), cows(1), d's(8), days(50), dies(2), does(116), dos(1), faze(1), five(182), gave(24), gays(10), give(100), goes(116), guys(71), has(194), have(980), hayes(4), haze(1), his(154), jazz(3), joe's(1), keys(2), knees(2), knows(31), laws(21), leave(37), live(137), lose(10), love(85), move(61), news(38), noise(2), p's(1), pays(2), phase(1), raise(19), rave(1), rise(1), rose(1), save(9), says(78), seas(1), sees(9), shave(1), shoes(13), shows(12), size(8), t's(1), taj(1), these(216), those(212), ties(1), toes(1), toys(2), twos(3), use(63), was(1634), wave(1), ways(41), whose(7), wise(10)

Voiceless (1823 tokens; 69 unique words), with individual counts: base(9), bash(1), bass(1), beef(1), biff(1), boss(3), bus(17), bush(8), calf(1), case(28), cash(3), chess(1), chief(7), choice(24), cuff(5), cuss(1), dose(1), face(23), fish(3), fuss(2), gas(6), geese(1), goose(1), gosh(28), guess(140), half(68), hash(1), house(134), joyce(2), juice(1), kiss(1), knife(1), las(1), laugh(2), lease(8), less(38), life(137), loose(1), los(1), mass(4), mess(6), mice(1), miss(13), moss(1), nice(86), niche(4), niece(3), pace(1), peace(7), piece(10), piss(1), push(12), race(6), rash(1), reese(1), rice(1), rough(12), rush(1), safe(5), this(707), tiff(1), toss(1), tough(21), vice(4), voice(7), wash(4), wife(47), wish(21), yes(122)

Appendix B

Competing Constraints Model

VC syllables. Constraints in this model are realized as Normally distributed probability densities. Probability decreases in either direction away from a maximum at the segment's preferred duration (μ); the rate of decrease is determined by the variance of the distribution, which is a proxy for segment elasticity. Probability densities function as gradient constraints under optimization of the joint probability. When preferred segment durations conflict with one another, the highest joint probability is achieved by violating lower-ranked constraints: i.e., shifting segments with higher elasticity further away from their preferred durations so that lower elasticity segments can remain closer to theirs. We assume that target syllable duration is determined by a combination of speaking rate and other prosodic factors, such as phrase-final lengthening (see, e.g., Byrd and Saltzman 2003).

The full set of constraints for the competing constraints model is given in (3), along with the parameter values used for the simulations. The mean values for the D, T and V distributions are roughly in line with observed values. The same is true of the relative variances: D has the smallest, then T, and V with the largest. Note, however, that the variance is a property of the constraint itself, and does not correspond to the actual variance of the category.

(3) $P(\frac{C}{V}) \sim \mathcal{N}(\mu = .3, \sigma = .1)$

$P(D) \sim \mathcal{N}(\mu = 50, \sigma = 15)$

$P(T) \sim \mathcal{N}(\mu = 50, \sigma = 60)$

$P(V) \sim \mathcal{N}(\mu = 100, \sigma = 70)$

$P(\frac{\sigma_T - \sigma}{\sigma_T}) \sim \mathcal{N}(\mu = 0, \sigma = .07)$

$\frac{V}{\sigma}$: V cannot be shorter than half the total syllable duration

The optimization function for this model, as a function of σ_T, and for any consonant, vowel pair (y, z), under the assumption of independence, is given as

$$p(y,z,\sigma_T) = p(\frac{C}{V} = \frac{y}{z}) \cdot p(C=y) \cdot p(V=z) \cdot p(\frac{\sigma_T - (y+z)}{\sigma_T}) \quad (4)$$

To reduce run time, the constraint ($\frac{V}{\sigma}$) is implemented by simply restricting the search space.[21] The dnorm() functions in R (v 1.4.1106) are used for the probability functions, with means and variances specified above.

The target syllable durations used for the corpus simulation were sampled from a Normal distribution with μ equal to 150 ms, and a σ of 200. These parameters were chosen to reflect the fact that almost all corpus vowel durations fell below the experimental cross-over point between voiceless and voiced percepts. Thus, sampled syllable durations are chosen to cluster in a range where there was little difference between the duration of the two obstruents. The same 1000 point sample of targets was used for both the voiced and voiceless distributions. The results are given in Section 4.1. See Figures 6 and 7.

Short Vowels: A short vowel, like a short consonant, can be specified with a lower elasticity. Because this is not a simple temporal compensation model, lower vowel elasticity does not automatically lead to significantly longer consonant durations. Figure B1 shows the result of reducing the variance of the vowel probability distribution ($\sigma = 40$: intermediate between that of the voiceless obstruent, and that of the voiced obstruent). The original (long vowel) model results are included for comparison. All other parameters remained the same, including the mean of the vowel distribution. The result is a smaller duration difference between the paired voiced/voiceless vowels, and shorter syllables over-all. Both obstruent types become slightly longer under these conditions, but the largest change is in how closely the target syllable duration is approximated. In this model, greater target undershoot results in a higher joint probability than increased lengthening of either consonant. Qualitatively, this behavior is consistent with the finding that the voicing effect is significantly reduced in preceding vowels that are inherently short (Umeda 1975; Crystal and House 1982; De Jong 2004). Note that the difference in duration between the

[21] Restricting the range effectively removes all values below a certain probability from consideration. Because this occurs before the joint probability is calculated, the restriction cannot be altered, making this constraint inviolable.

obstruents themselves can, in principle, still grow quite large. Because very few studies on the voicing effect report final obstruent durations, it remains to be seen whether this prediction is borne out.

[H]

(a) *Original model (long vowel)* (b) *Short vowel*

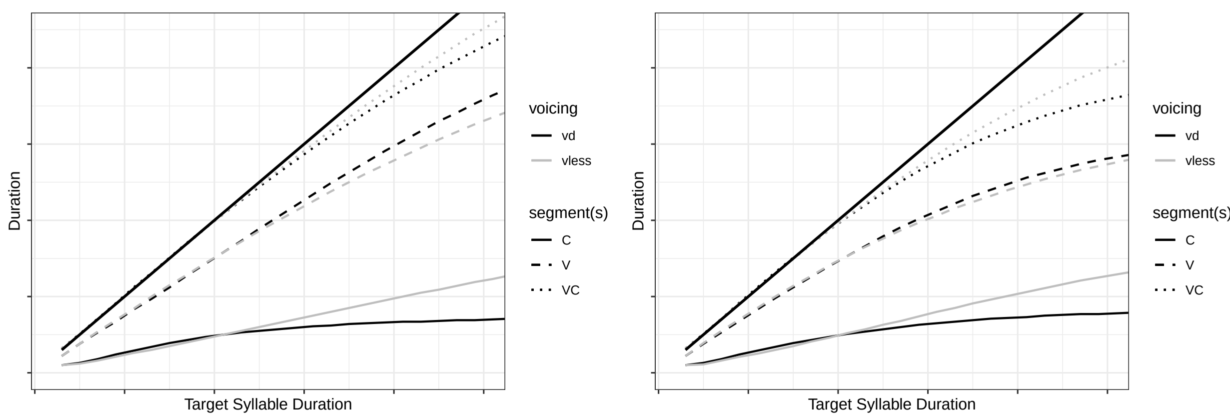


Figure B1
Competing Constraints Model

CV syllables. Resistance to lengthening under slowed speaking rate is also observed for voiced stops in initial position. Furthermore, in the handful of studies that vary vowel, rather than syllable, duration the results are qualitatively similar to what is found in medial and final position: perception of voicing switches based on vowel duration, but only for longer vowels. Viswanathan et al. (2019) report a significant difference in voicing perception between vowels of 175 and 225 ms., but no difference between longer vowels, at 225 and 275 ms., and no difference between shorter vowels, at 125, 150 and 175 ms. Toscano and McMurray (2015) find a significant difference in response rate across the same boundary, between vowels of 189 ms. and those of 377 ms.

VOT ranges for stops in initial position are comparable to what we see for closure durations in final position in the corpus and in the production study (Section 3): from roughly

50-150 ms for the voiceless stop, and 10-50 ms for the voiced (Allen and Miller, 1999; Pind, 1995; Miller and Baer, 1983; Miller et al., 1986). However, post-voiced vowels were only 10-19% longer than post-voiceless, compared to 30-40% for pre-voiced versus pre-voiceless. As with medial position, we attribute some of the smaller effect size to the lack of a final lengthening effect.[22]

Additionally, pre-vocalic and post-vocalic consonantal gestures are phased differently in English; singleton consonants in onset are activated at the same time as the vowel, whereas coda consonants are activated at the offset of the vowel (e.g., Browman and Goldstein 1988). See Section 5. The onset-vowel phasing relationship is also less variable (e.g., Selkirk 1982), resisting adjustments that would shift the two segments apart. As a result, part of the vowel is consistently masked by the consonant, and thus acoustically shorter than a bare vowel. Inherently longer consonants will lead to greater masking than inherently shorter consonants. Thus, it is predicted that the vowel following a voiceless obstruent will be acoustically shorter than a vowel following a voiced one, but only in the range where voiceless obstruents are longer than their voiced counterparts.

By making two changes to the VC model, the predicted behavior of the onset "voicing" effect can be reasonably well-captured. The target-matching constraint is altered to apply only to the vowel, and not to the onset of the CV syllable. This assumption is necessary to produce a different outcome from the VC case. But it is also based on the fact that onsets do not typically take part in prosodic phenomena, being irrelevant to the calculation of syllable weight, for example (e.g., Hyman 2019). Changing the relevant unit from syllable to rhyme in Eq. 4 will cover both the VC and CV cases. The rhyme in the VC case is calculated by adding the durations

---

[22] While lengthening occurs preceding, as well as following, prosodic boundaries, the effects are not the same. The consonant immediately following a prosodic boundary shows lengthening, but the vowel following that consonant typically does not (e.g., Fougeron and Keating, 1997; Cho and Keating, 2009; Kim and Cho, 2012). Byrd et al. (2005) find that coda consonants show a larger difference in duration when compared across medial and boundary position than onsets. Interestingly, the difference seems to lie in the fact that more of the coda gesture is lengthened. For both codas and onsets, the portion of the gesture closest to the boundary is lengthened the most – for codas the release portion, and for onsets, the constriction portion. However, the constriction portion for the coda consonant is also lengthened to a lesser degree, while the release for the onset is not (or at least, not consistently).

of the consonant and vowel (assuming no overlap).[23] The rhyme in the CV case is calculated from the articulatory duration of the vowel; the acoustic duration of the vowel is given by subtracting consonant duration from articulatory duration (assuming full overlap).

The second change is a reduction in the variance of the C/V constraint (by 60%, measured with respect to the articulatory duration of the vowel). By making the variance smaller, the outcome becomes more strongly biased towards the preferred C/V value than it is towards perfect rhyme duration matching.[24] This is also consistent with the lower variability in phasing between onset and nucleus, versus nucleus and coda. All other model parameters remain the same.

These changes alter the model behavior in the desired ways. See Figure B2. The VC model (long vowel) is included for comparison. Even for the small set of constraints used here, the interactions are complex. However, we can broadly outline the effects of changing the parameters in the way described. In the coda model, duration differences between the obstruents arise because of the smaller variance of the voiced obstruent duration constraint. The interaction of this constraint with the targeted rhyme duration constraint gives rise to the complementary difference in preceding vowel duration. In the onset model, the obstruent duration constraints remain the same, causing lengthening of the voiced obstruent to be more costly (reduce probability more), than lengthening of the voiceless obstruent. However, onset duration is not relevant to the target rhyme constraint in the CV case, so there is no interaction. The only pressure to lengthen the consonants comes from the C/V constraint. Therefore, consonants only lengthen in order to achieve the best possible C/V ratio. The consonant durations, however, do not differ much from the previous models. It is the vowel duration that is most strongly affected by the re-weighting of the C/V constraint. The articulatory vowel faces pressure to lengthen, but undershoots the target more than in the original VC model, due to the greater influence of the C/V constraint.

---

[23] If the operative unit is the rhyme, and duration is specified at that level, then this model can also account quite simply for the finding that vowels in open syllables are typically longer than vowels in closed syllables. For the same target rhyme duration, and a highly weighted target matching constraint, the same duration is distributed over two segments in the (C)VC case, and only one in the (C)V case.

[24] A smaller variance will increase resistance both to C/V values that are too large, as well as those that are too small.

[H]

(a) *Coda model (original)*

(b) *Onset Model. V is acoustic vowel length. CV is articulatory vowel length.*

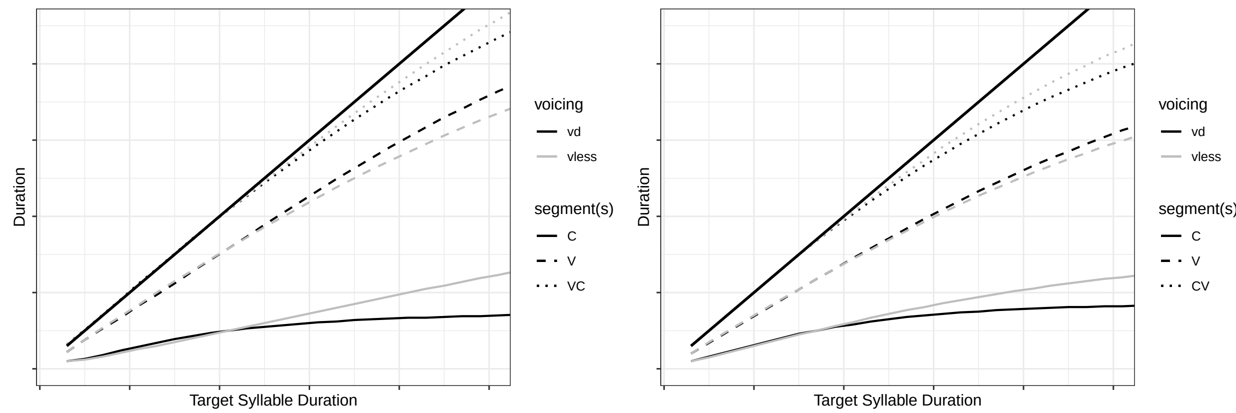


Figure B2
Competing Constraints Model

The general behavior of the onset model looks very similar to the coda model with the inherently short vowel. However, the mechanism is quite different; the relevant variable is a ratio (C/V), rather than a sum ($Rhyme = V + C$). Therefore, the relationship between vowel and consonant duration is not negatively correlated, but positively correlated: the articulatory vowel is *shorter* because the tautosyllabic consonant is shorter. Vowels following voiced consonants are therefore shorter than vowels following voiceless consonants (CV in Fig. B2). However, the in-phase timing between onset and nucleus means that the articulatory vowel will be masked to a greater degree by the longer (voiceless) consonant. In this case the effect of masking is slightly larger than the C/V effect. Therefore, the net result is a slightly longer post-voiced than post-voiceless acoustic vowel. The voicing effect is smaller, but it shows the same dependence on total duration as the other models. These outcomes are consistent with the literature.

Appendix C

Corpus: Statistical Models

7.1 Complex vowel duration models

The Tanner et al. (2019) model formula is given in (5). Because this model would not converge for our data, it was simplified. In order to compare the voicing to the consonant duration model with the same random effect structure we had to simplify further to (6) and (7).

(5) Vowel Duration ~ Local Speaking Rate + Global Speaking Rate + Frequency + Vowel Class + Vowel Height + Voicing + Stop Type + Word Class + Voicing:(Frequency + Vowel Class + V Height + Stop Type + Word Class) + (Frequency + Vowel Class + Vowel Height + Stop Type + Voicing + Word Class + Voicing: Stop Type | Speaker) + (Local Speaking Rate + Global Speaking Rate | Word) + (1 | Vowel Quality)

(6) Vowel Duration ~ Vowel Height + Vowel Class + Voicing + Speaking Rate Deviation + Obstruent Type + Frequency + Voicing:(Speaking Rate Deviation + Obstruent Type + Frequency) + (Voicing + Obstruent Type + Frequency|Speaker) + (1|Word) + (1|Vowel Quality)

(7) Vowel Duration ~ Vowel Height + Vowel Class + Consonant Duration + Speaking Rate Deviation + Obstruent Type + Frequency + Consonant Duration:(Speaking Rate Deviation + Obstruent Type + Frequency) + (Consonant Duration + Obstruent Type + Frequency|Speaker) + (1|Word) + (1|Vowel Quality)

The results for phrase-final tokens are given in Table C1, for voicing, and Table C2, for consonant duration. The interaction between speaking rate and voicing, and between frequency and voicing are both significant, although voicing fails to reach significance on its own. In model 2, consonant duration is significant, as is obstruent type, and there is a positive interaction between consonant duration and obstruent type. The negative effect of consonant duration is enhanced for fricatives.

As with the simple model, the consonant duration model provides a better fit to the data: AIC score of 1320 vs 2019.

[H]
Table C1
Utterance Final: Interactions Voicing Model.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(intercept)	4.81	0.05	25.5	107	< 2e-16
Vowel Height	0.17	0.04	10.6	4.68	7.4e-4
Vowel Class	0.11	0.03	10.8	3.44	5.6e-3
Voicing	-0.01	0.02	124	-0.59	0.56
Speaking Rate	0.67	0.01	3228	45.6	< 2e-16
Obstruent Type	0.02	0.02	136	0.97	.33
Frequency	-0.10	0.03	115	-3.44	8.1e-4
Voicing: Rate	0.06	0.01	3109	4.19	2.9e-05
Voicing: Obs. Type	-0.01	0.01	229	-0.99	0.32
Voicing: Frequency	-0.05	0.03	230	-2.17	0.03

[H]
Table C2
Utterance Final: Interactions Duration Model

	Estimate	Std. Error	df	t value	Pr(>|t|)
(intercept)	4.81	0.04	39.0	109	< 2e-16
Vowel Height	0.19	0.03	9.29	9.29	8.7e-5
Vowel Class	0.13	0.03	9.77	4.96	6.1e-4
Speaking Rate	1.01	0.02	3190	56.4	< 2e-16
Consonant Duration	-0.52	0.03	76.6	-20.4	< 2e-16
Obstruent Type	-0.07	0.02	159	-3.93	1.3e-4
Frequency	-0.07	0.03	146	-2.43	0.02
Consonant Duration: Rate	-0.005	0.02	2995	-0.37	0.79
Consonant Duration: Obs. Type	0.05	0.01	2681	3.67	2.5e-4
Consonant Duration: Frequency	-0.02	0.03	2297	-0.65	0.52

It is not entirely clear why we fail to find a voicing effect where Tanner et al. (2019) did find one. It could be because we included very short tokens, or because we had a smaller sample of data. However, removing tokens under 50 ms did not appreciably change the results, and a model of non-phrase-final tokens did not produce a significant voicing effect either. The local speaking rate measure used by Tanner et al. (2019) is a count of the number of segments within an inter-pause interval containing the target word, therefore it is much less local than the one we use. This type of measure also does poorly in estimating rate for pre-pausal tokens that are subject to final lengthening, and at the other end of the continuum, with the very shortest tokens. These two ranges show non-linear behavior for this type of speaking rate measure, which is presumably why linear regression models do not capture them well (see Gahl, 2009). If less of the over-all variance is attributed to speaking rate, this might account for the difference between our results.

7.2 Consonant duration model

Consonant Duration ~ Voicing + Speaking Rate + Obstruent Type + Frequency + Voicing:(Speaking Rate + Frequency + Obstruent Type) + (1 | Speaker) + (1 | Word)