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

Rebecca L. Morley and Bridget J. Smith

The Ohio State University

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

The terms “vowel lengthening” and “voicing effect” are used to refer to the highly-replicated empirical finding that vowels preceding voiced obstruents tend to be 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). The bulk of the literature focuses on varieties of English, but voicing effects have been documented in a number of different languages. In some, such as Arabic and Catalan, the effect appears to be quite small, while in other languages, such as French, Norwegian, and Korean, larger differences have been found (House and Fairbanks 1953, Abdelli-Beruh 2004, Hillenbrand et al. 1984, Mack 1982, Laeufer 1992, Elert 1965, Kulikov 2012, Cuartero Torres 2002, De Jong and Zawaydeh 2002, Chen 1970). It is generally agreed that English (in many of its varieties) exhibits one of the strongest voicing effects, where pre-voiced vowels can be from 30% 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). The fact that other cues to voicing have been shown to be unnecessary in such experiments, or, in fact, to be secondary to vowel duration, has led to the claim that vowel duration is not only a sufficient cue to the voicing contrast on final obstruents in English, but the primary cue (e.g., Raphael 1972, Luce and Charles-Luce 1985). Because stops are often unreleased in word-final position, it has been argued 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).

Despite the large amount of research on the phenomenon, the underlying source of the voicing effect remains an open question. There is little consensus on what acoustic or articulatory properties give rise to the observed duration differences. Nor is the effect even consistently described as lengthening, but sometimes as shortening before voiceless consonants, or “pre-fortis clipping” (Gimson 1970, Wells 1982). We hypothesize that this uncertainty persists, in large part, due to a widespread over-simplification of the empirical facts. In the first place, it has been known for some time that lengthening does not occur in all environments, or for all vowels (e.g., Peterson and Lehiste 1960, Umeda 1975, Hogan and Rozsypal 1980, Crystal and House 1982, De Jong 2004). Neither does preceding vowel duration robustly cue the voicing contrast at all speaking rates and for all types of speech (Umeda 1975, Port 1976, Crystal and House 1988, Smith 2002, Ko 2018). Other cues, such as voicing itself, or a period of aspiration following a stop release, may supersede vowel duration in perception (e.g., O’Kane, 1978, Raphael, 1981, Wardrip-Fruin, 1982, Revoile et al., 1982, Repp and Williams, 1985). Finally, lengthening can occur even when voiced obstruents are not actually voiced (i.e., in the absence of vocal fold vibration) (e.g., Sharf 1964, Walsh and Parker 1981, Keating 1984).[1]

In this paper we will argue that durational differences are not the direct result of voicing at all (whether phonetic or phonological), but of intrinsic segment elasticity, which results in voiced obstruents being shorter on average than voiceless obstruents. The inverse correlation between differences in obstruent duration and differences in preceding vowel duration led both Kozhevnikov and Chistovich (1965) and Catford (1977) to propose that the former is what gives rise to the latter. However, temporal compensation as an explanation for the voicing effect has

[1] The apparent paradox involved in describing [+voice] segments as voiceless, or devoiced, has been traditionally resolved by differentiating between a phonetic voicing feature, and a phonological voicing feature. The phonological feature is to be thought of as a completely abstract label which can be transformed through various rules to either a phonetically voiced or phonetically voiceless realization. The question of the best way to characterize laryngeal contrasts phonologically will be taken up in Section 6.

been explicitly considered and rejected on a number of separate occasions. Chen (1970), for example, found that syllable duration was not uniform across CVC and CVCC words, indicating that vowels in the latter type of syllable were not shortening to compensate for the added duration of the second coda consonant. This led him to rule out compensatory mechanisms altogether. Braunschweiler (1997) reached a similar conclusion based on the large duration differences between VC sequences containing a short versus a long vowel. Keating (1985) also rejected a compensation account, based on Polish data in which closure duration, but not preceding vowel duration, varied across voiced-, and voiceless-final syllables. In English, CVC syllables tend to be longer when closed with a voiced obstruent, than with a voiceless, which can also be taken as evidence against a compensation account (e.g., Jacewicz et al. 2009, Luce and Charles-Luce 1985). Such arguments are based on the assumption that temporal compensation arises from a pressure to keep syllable length uniform (isochrony), meaning that compensation should be total, or near total.

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, 1957a, 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. As a result, the effect sizes were quite small, with duration differences for both vowels and stop closures ranging between 8 and 35 ms. In a variable-rate production study, at much longer durations, we also find a close correspondence between vowel and obstruent duration differences. However, the task itself is likely to encourage explicitly compensatory behavior. Following work in Articulatory Phonology on other, apparently compensatory, phenomena (e.g., O'Dell and Nieminen 1999, Nam and Saltzman 2003), we adopt the position that apparent compensation is emergent from the interaction of conflicting gestural timing constraints.

In our model, individual segments resist pressures at the syllable level to lengthen or shorten from their characteristic, or preferred, durations. At short durations/fast speeds, voiced and voiceless obstruents are similar in duration. At longer durations/slower speeds, vowel differences mirror consonant differences, not because all syllables must be the same duration, but because voiced obstruents resist lengthening much more than vowels and voiceless obstruents do (see Cambier-Langeveld (2000) and Miller 1981 on inherent segment elasticity). The longer the syllable gets, the less expandable each segment grows. However, as long as the syllable continues to lengthen, those segments whose expandability decreases the most rapidly will account for less and less of the total syllable duration, while the more expandable segments take on an increasingly larger proportion. We call this the Expandability Hypothesis.

(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

Relative resistance determines the distribution of duration across the syllable

We will show that the Expandibility Hypothesis parsimoniously accounts for data on the voicing effect from both production and perception, predicting that vowel duration differences should only be seen when there is a complimentary difference in consonant duration, and that the size of the effect should increase with increasing duration. The Expandability Hypothesis also unifies a number of other duration-based phenomena under a single account, including other instances of apparent temporal compensation,

The paper is organized as follows. In Section 1 we provide a critical review of other explanations of the voicing effect. Section 2 contains a corpus study on American English, the results of which we model in Section 3 under the Expandability Hypothesis. In Section 4, predictions of this hypothesis are tested using a variable-rate production task. The Expandability Hypothesis is elaborated in Section 5, where we provide an explanation 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 Explanations for the Voicing Effect

The majority of the literature on the voicing effect seems to assume an articulatory source without further discussion (e.g., Raphael 1975, House 1961, Ohala 1983). Belasco (1958), however, speculates that there is a trade-off in the force of production between the vowel and coda consonant of a syllable. When the consonant requires more energy, or effort, the vowel is altered to require less, and vice versa. Thus, voiceless stops, involving forceful release and aspiration, condition shorter vowels, which require less energy. Similarly, Delattre (1962) argues that anticipation of a effortful articulation should shorten the preceding vowel. However, Moreton (2004) and Schwartz (2010) argue essentially the opposite: that it is the spread of "fortisness", or "hyper-articulation" that shortens the preceding vowel.[2] It has also been claimed that careful, and therefore slower, movements of the vocal cords are required to avoid spontaneous voicing under reduced pressure (Halle and Stevens 1967); or that the transition from vowel to voiceless obstruent is more rapid than the transition to voiced (Chen 1970); or that the glottal opening gesture tends to occur a bit earlier for voiceless final consonants to ensure that there is no residual voicing on the consonant (Klatt 1976). However, clear evidence of differences in energy, effort, or precision between voiced and voiceless obstruents has not been forthcoming. Additionally, as Lisker (1974) points out, the cause and the effect for many articulation-based explanations cannot be assumed. Voiceless stops may involve earlier glottal opening, and a more rapid transition from the vowel, but those facts only explain the shorter duration of the vowel if vowel shortening is an unavoidable consequence of those properties of articulation. It could as easily be said that such articulatory properties are explained as the consequence of producing the desired voiceless stop.

On the auditory side, Lisker (1957a) suggests that longer vowels occurring with shorter voiced obstruents, and shorter vowels with longer voiceless obstruents, is an enhancement effect, reinforcing the length differences of the obstruents, and thus the voicing contrast (see also Jessen

[2] In Moreton (2004) this also serves to explain why such vowels are more phonetically dispersed, or peripheralized

2001, and Kluender et al. 1988). Javkin (1977) posits that vowels are consistently perceived as longer before voiced than voiceless consonants because listeners mis-attribute the glottal pulsing at the beginning of the consonant 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.

The above explanations encounter more problems when the voicing effect is investigated more closely. Those of 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). 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 also tend to be monosyllabic, which entails that the target vowel receives primary stress. When some or all of these factors are varied, the voicing effect can disappear.

2 A Corpus Study

Data from the Buckeye Corpus (Pitt et al. 2007) were analyzed in order to determine the properties of the voicing effect in actual usage. The corpus consists of segmented and transcribed sound files from several different speakers, collected during individual interviews, each lasting approximately one hour. From this 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). 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 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.

It is expected that speech collected from different individuals, uncontrolled for any linguistic variables, will show a high degree of variability. Conversational styles of speech are also 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). Nevertheless, for listeners to use a given feature in discriminating contrastive sound units, there must be some way for that feature to be extracted from the actual speech signal. To get a sense for the size of the voicing effect relative to other factors affecting vowel duration, we include a number of different visualizations of the data to accompany our statistical analysis.

In Figure 1 vowel durations for the entire set of CVC word tokens 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 (almost 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 (>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.

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 actually a diphthong in some dialects, thus it is grouped with other inherently long (tense) vowels. Reduced or absent voicing effects have been reported for both unstressed and lax vowels (Umeda 1975, Crystal and House 1982, De Jong 2004). 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.

- SPEAKING RATE DEVIATION: The z-scored average difference between expected and observed word duration was used as a proxy for rate difference (see Gahl et al. 2012, Priva 2017). Expected word duration was taken to be the sum of the expected durations of the individual segments within the word. Mean segment duration over all word tokens was used as the expected value, calculated separately for each speaker. 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. The following tags were used to identify a boundary: SIL (silence), E_TRANS, 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: 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: voicing 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.

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

Like vowels, consonants possess different inherent lengths. Furthermore, stops may be reduced in certain contexts to flaps, which consist of a very brief tongue tip gesture against the roof of the mouth. Because the durations of each have their own distributional characteristics, and may have different effects on the preceding vowel, each manner (fricative, stop, and flap) was analyzed in a separate statistical model. For each model, the factors listed above were included as main effects. All factors were sum-coded so that each individual factor was assessed at the mean value of all other factors.

Continuous numerical variables were log-transformed, where appropriate, and mean-centered to approximate a normal distribution with a mean of zero. Random intercepts for word and speaker were included in all models. Random slopes were added for all factors when doing so significantly improved model fit. 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 very skewed distribution of the data, it was not possible to used paired data in analyzing the voicing effect (see Section 2.2).

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. Interactions that did not reach a significant estimated t-value were individually removed from the model, and goodness-of-fit model comparisons confirmed that including these factors did not improve model fit. Three-way interactions were avoided for reasons of interpretability as well as model convergence.

The results were similar for the full stop and fricative models. Stops and fricatives showed the same significant effects, and similar relative effect sizes, except for the final interaction term. Because of these similarities, and for reasons of space, only the results for stops are reported here.

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 Short vowels is significantly different from the global vowel duration average, calculated over both Long and Short vowels.

As expected, there was a significant main effect of speaking rate. See Table 1. Longer vowel durations were found at slower speaking rates. However, the negative effect of consonant duration indicates that, for words of average speaking rate and average frequency, vowels were longer when final consonants were shorter. Similarly, word frequency had a significant negative effect on vowel duration, such that more frequent words had shorter vowel durations. As predicted, Short vowels were also shorter than Long vowels, and vowels in pre-pausal words were considerably longer than in non-prepausal. However, both phonetic and phonemic voicing actually had a negative effect on vowel duration, the opposite of what is predicted by the voicing effect.

The interaction of speaking rate and phonemic voicing, however, shows a positive adjustment, meaning that the negative correlation is lessened at slower speaking rates, and thus more in line with the predicted directionality. A similar interpretation can be given to the interaction of frequency and phonemic voicing. The negative effect of frequency (lower frequency = longer vowel duration) is larger in voiced, than voiceless, tokens. Thus, the negative effect of voicing (voiceless tokens longer than voiced) is reduced at lower frequencies. A positive interaction of consonant duration and speaking rate also indicates that at slower speaking rates, the inverse effect of consonant duration is smaller. This is likely just because there is a positive correlation between speaking rate and consonant duration: all durations get longer with slowed speaking rate. The interaction between speaking rate and vowel class shows a positive adjustment to short vowels at slower speaking rates. Even though they may be inherently shorter, Short vowels still lengthen as speaking rate slows, and proportionally more so than Long vowels, reducing the effect of vowel class on vowel duration. Finally, the significant interaction between phonemic voicing and consonant duration (for stops but not fricatives – which show more variability over-all) indicates that the negative correlation of consonant duration with vowel duration is larger when the consonant is the voiced member of the contrast. This may simply be because there is less variability in the duration of voiced stops.[3]

2.1 A note on collinearity and speaking rate

The syllable is generally taken to be the relevant unit of speech timing in English, based in part on the finding that segment durations show high variability, while larger units, such as the syllable, foot and word, show much less variation (e.g., Allen 1975). All syllables, however, do not behave the same. There is more variability in syllables containing an inherently long vowel, than in those containing an inherently short vowel. Not only are unstressed syllables much shorter than stressed syllables, they respond differently to changes in speaking rate. For this reason,

[3] Beguš (2017) also finds an interaction between voicing and closure duration, which he attributes to the effect of the abstract laryngeal feature [+voice].

Peterson and Lehiste (1960) adopted the inter-stress interval (time elapsed between stressed vowels) as the operative measure of rate. However, the most common way to measure speaking rate, syllables per unit time, ignores the differences in syllable type. In cases where reduction has led to vowel deletion, the count is usually taken over the number of syllables in the citation form of the word.

The choice of the interval of speech over which speaking rate is to be estimated varies across different studies. Often the interval corresponds to an utterance or a phrasal unit which contains the word of interest (Pickett and Decker 1960, Quené 2008, Gahl et al. 2012, Bell et al. 2009). For continuous speech, such an interval is usually equated with the interval between two pauses of some minimal duration. It was originally assumed that larger intervals, spanning multiple conversational "turns", were better for accuracy in estimates of speaking rate. However, it has since been found that speaking rate is likely to vary considerably within such an interval (Miller et al. 1984). In fact, speaking rate may vary significantly over a much shorter time scale. In experiments in which speech rate is deliberately varied across a given utterance, the material closest to the target word has been shown to have the largest effect on perceived speaking rate; this is the case when the manipulated material is the immediately preceding word, the immediately preceding segment, or even the transition rate from the preceding segment (Verbrugge and Isenberg 1978, Summerfield and Haggard 1972, Ainsworth 1972, Lindblom and Studdert-Kennedy 1967, Summerfield 1981). Thus, the durational properties of the target word (or syllable) can be used as a measure of speaking rate. In practice, because vowels generally show the largest changes under changes in speaking rate (e.g., Gay 1978), and vowel duration tends to be the strongest perceptual cue to speaking rate (Crystal and House 1982, Summerfield 1981, Port and Dalby 1982, Ainsworth 1972, Lindblom and Studdert-Kennedy 1967), vowel duration is often used as a proxy for speaking rate. This is clearly problematic when the vowel duration is assumed to be a product not only of speaking rate, but of the phonological properties of the segments within the syllable as well.

In studies that use speech corpora, the number of syllables contained within the inter-pause interval that also contains the target word is commonly used to estimate speaking rate. However, there is often more than one measure of speaking rate included in such analyses. For example, in their own analysis of the Buckeye Corpus, Gahl et al. (2012) used three rate measures. They estimated rate in the post-pausal interval before the target word, as well the pre-pausal interval after the target word. In addition, they included a measure of expected, or baseline, word duration. This was calculated by summing the average segment durations within the word – where averages were taken over the entire corpus. This measure was meant to unconfound differences in duration due to inherent segment characteristics from those due to speaking rate. Tanner et al. (2019), who also analyzed the Buckeye Corpus, used two different speech rates: a global average for each speaker, and a local speech rate given by the deviation from the global mean (interestingly, only local rate was found to affect the size of the voicing effect).

The reason for the use of multiple measures, and the variability across studies, is twofold. In the first place, all estimates of rate are actually estimates of duration. It is impossible to avoid this confound when using acoustic data that has not been generated in explicitly rate-controlled contexts. Therefore, any study of duration effects must seek to make speaking rate estimates as independent of local duration as possible. In the second place, it is simply not known how much speaking rate can be expected to vary. If the target word is phrase-medial, then relatively larger intervals may provide appropriate estimates, but it is known that phrase-final words are not well modeled by the inter-pause intervals, whether they include the target word or not. These are among the longest words in any given corpus. A similar problem arises with words on the very shortest end of the continuum. These two ranges show non-linear behavior with respect to the typical speaking rate measures, which is presumably why linear regression models do not capture them well (see Gahl, 2009).

There seems to be an emerging trend to use more complex methods for estimating speaking rate, especially those based on comparison between observed and expected (baseline) values (See Priva 2010, 2017). This has motivated our choice of speaking rate calculation. However, practical considerations were the largest factor in this decision. To avoid the issues with the longest tokens, other models have simply excluded them. However, such tokens were the focus of our study. And we found that simpler metrics failed to show the expected dependence between duration and speaking rate. Therefore, it was decided to use the estimation procedure described above. The basic procedure follows Gahl et al. (2012), Tanner et al. (2019), and Priva (2017) in comparing an expected value against the value that is actually observed. We explicitly model phrase position, vowel identity, coda voicing, word frequency, and coda duration, thus any difference between expected and observed duration that is not highly correlated with one or more of these factors we attribute to a departure from the average speaking rate. It is true that any other sources of duration difference, such as place of articulation, manner, or voicing of the word-initial consonant, will also be folded into this measure. However, the effects of such factors are likely to be small, and statistically unrecoverable in the presence of much larger effects. Therefore, what remains is a reasonable estimation of the effect of speaking rate.

2.2 The Voicing Effect

The lack of a main effect for voicing in the predicted direction indicates that this cue is not extractable from the corpus in the aggregate, even when phrase-final position and other factors are partialled out. Part of the reason for this is the unbalanced nature of the data (to which we attribute a spurious negative correlation of voicing and vowel duration). The skew across multiple variables is notable, not just numbers of voiceless stop versus voiced stop tokens, but the distribution over vowels, frequency and speaker. To the degree that this type of speech data is typical, however, the listener/learner faces a similar analysis problem.

However, the interactions of speaking rate deviation with voicing, and frequency with voicing, are suggestive. The directionality is consistent with a voicing effect that is dependent on absolute duration. Slower speaking rates and lower frequencies both result in longer words tokens. In such contexts pre-voiced vowels start to “catch up” to pre-voiceless vowels in duration. This interpretation is also consistent with the reported findings that the voicing effect is smaller, or missing altogether, in phrase-medial position, for lax vowels, for unstressed vowels, and in polysyllabic words. Each of these contexts effectively act to shorten word tokens. To test this hypothesis further, we extracted the subset of data corresponding to the upper 50% quartile of each model variable. The log-normed vowel durations are plotted in Figure 2, separately for stops and fricatives: content words of below median word frequency, containing only the Long class of vowels, produced at slower than average speaking rates, in pre-pausal contexts.

Although still very small, the duration difference now appears to go in the right direction. Statistical analysis confirms that there is a significant positive effect of voicing. The full model is given in Table 2. As before, separate linear mixed effects models of vowel duration for words ending in fricatives and stops were constructed.[4] Again, due to highly similar results between fricatives and stops, only the stops are reported. The model is simplified because there were fewer data points, and less variability in the factors. Word frequency and speaking rate were no longer significant. This is not unexpected, given that the lower half of both distributions were excluded. The large effect of phrase-finality is also likely to wash out any remaining small effects. No interactions reached significance.

Because all word tokens are longer than average, the effect of consonant duration on vowel duration now goes in the opposite direction. Longer consonants predict longer vowels because both are subject to the same lengthening effect of pre-pausal position. However, there is now a significant main effect of phonemic voicing in the predicted direction. This effect, we argue, is driven by a difference in consonant duration. Shorter consonants predict longer vowels, as can be seen when the full set of data is analyzed, with a comprehensive range of vowel and consonant durations. Pre-pausal lengthening swamps the inverse correlation of consonant and vowel duration in the smaller set of data, but the voicing variable allows it to be partially recovered by grouping shorter things (voiced obstruents) separately from longer things (voiceless obstruents). We thus hypothesize that the positive correlation of voicing with vowel duration derives from the negative correlation of voicing with consonant duration. This correlation reaches significance in this model, and not the previous one, because voicing is only a good predictor of obstruent duration when durations are longer than average.

[4] It is not expected that flaps will appear in this subset of data, as they are inherently shorter phones

2.3 Obstruent duration

The distribution of obstruent duration over voiced versus voiceless segments looks strikingly similar to the vowel duration distribution. 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. Figure 3 is the analogue to Figure 1, showing the raw consonant durations as both counts and probability densities. This figure also shows that voiced fricatives are more frequent than voiceless (voiceless fricatives comprise only 33% of the total) – the opposite of the distribution of the stops (voiced stops comprise only 22%). Furthermore, stops outnumber fricatives, at 55% of the total obstruent distribution.

Extracting the longest subset of the data reveals a similar trend to that observed for vowel duration. Slower speaking rates, for Long vowels, in phrase-final contexts, in low-frequency words, result in more separation between the voiced and voiceless distributions. See Figure 4. We confirm that voicing is significantly negatively correlated with obstruent duration ($p < .001$) in a separate model, with obstruent duration as the dependent variable, and fixed effects of speaking rate, consonant voicing, vowel duration, and phrasal position. It is worth noting that, even at average speaking rates, consonant voicing is a significant predictor of consonant duration. The correlation of voicing with consonant duration thus appears to be more consistent than the correlation of voicing with vowel duration. 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.

2.4 Summary & Discussion Of Results

The corpus results show that, in conversational speech, we do not see the expected effect of voicing on vowel duration at average speaking rates, for average word frequencies, or phrase-internally. We only begin to see an effect emerge at slower speaking rates, in lower frequency words, for longer vowels, and in phrase-final position. This dependence is seen in interaction terms for the full model, and in a separate analysis of phrase-final tokens. Tanner et al.

(2019), although they restricted their analysis to utterance-final tokens, also found that the voicing effect was smaller in faster words, at higher frequencies, and for inherently shorter vowels (high versus non-high) in the Buckeye Corpus. They interpret these results as evidence of reduction, or masking, of an inherent voicing effect. Our analysis, on the other hand, shows that consonant duration is a more consistent predictor of vowel duration than phonemic or phonetic voicing. The contexts in which voicing emerged as significant (in the right direction) can all be explained as effects of consonant duration, a factor that is correlated with speaking rate, phrase position, frequency and, critically, the voice feature. Thus, these results support our hypothesis that the voicing effect is fundamentally a duration effect, and one which is only significant at sufficiently long durations.

3 The Expandability Hypothesis: Modeling the Corpus Data

Pre-voiced vowels produced in laboratory settings can be up to 50% longer than pre-voiceless vowels, with reported durations in the range of 175 to 300 milliseconds (Peterson and Lehiste 1960, Mack 1982, House 1961, Luce and Charles-Luce 1985, Umeda 1975). 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 only characteristically longer vowels, only 13% of such tokens fall in this range. Median vowel duration for the subset of tokens plotted in Figure 2 is 200 ms, while 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.

Obstruent duration distributions show a similar pattern. Luce and Charles-Luce (1985) find that closure duration fails to reliably differentiate the voicing contrast on stops in most environments (e.g., across sentence position, place of articulation, and preceding vowel quality). However, in sentence-final position (pooling two tense vowels, and three places of articulation), they report voiceless closure durations on average 25% longer than voiced. Chen (1970) reports closure durations that are 50% longer, for monosyllables spoken in isolation. Absolute duration values for voiceless stop closures in these studies range from 95-140 milliseconds. 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.

The discrepancy in both vowel and consonant durations implies that laboratory speech falls in the very upper range of conversational speech (cf. Gahl et al. 2012). A strong dependence on absolute duration explains why no voicing effect was found for the full set of CVC tokens measured. Even in the lab, inherently shorter tokens show considerably reduced voicing effects. Not only is the vowel duration difference smaller for phrase-medial position (versus phrase-final), syllable-medial position (versus syllable-final), polysyllabic words (versus monosyllabic), lax vowels (versus tense), unstressed vowels (versus stressed vowels), high word frequency (versus low word frequency), and fast speaking rates (versus slow speaking rates), but differences between the associated voiced and voiceless obstruents are also small or nonexistent. (e.g., Umeda 1975, Port 1976, Crystal and House 1988, Smith 2002, Ko 2018, Hogan and Rozsypal 1980).

3.1 A pure compensation model

The temporal compensation hypothesis associated with the voicing effect is almost always a perfect compensation hypothesis (see Chen 1970, Keating 1985, Port and Dalby 1982). Differences in vowel duration are expected to be approximately equal in magnitude to differences in consonant duration for a given minimal pair. This type of model is implemented in Campbell (1992). Duration is specified at the syllable level, and distributed over the segments within the syllable according to their relative elasticity. Campbell's hypothesis is that patterns of variation at the segment level can be derived from only two parameters: the inherent elasticity of the segment (which is fixed), and what we will call the expansion coefficient (ε), which varies as a function of target syllable duration (see also Campbell and Isard, 1991 and Campbell, 1990). The function for calculating the expansion coefficient for a given syllable, σ_k, is given in Equation (2): the solution is the value that, when distributed to each segment in the syllable according to their specific elasticities (κ_i), will result in the necessary total duration change from the underlying syllable duration ($\bar{\sigma}_k$), to the target syllable duration (σ_T).

$$\varepsilon_k(\sigma_T) = \frac{\sigma_T - \bar{\sigma}_k}{\sum\limits_i \kappa_i} \tag{2}$$

It has been observed that longer segments generally show more variability in their duration (e.g., Lehiste, 1972). That longer segments should have larger variance, rather than just larger means, can be explained by the asymmetries we have already seen in the Buckeye Corpus (also reported by Crystal and House 1988). In this corpus, and presumably others containing the same speech style, the duration distributions for longer and shorter segments are very similar, varying primarily in their upper limits. Longer segments thus contain a superset of the duration range of shorter segments, resulting in larger variance. If variability is a direct indicator of elasticity, then it follows that most of the modifications in syllable duration will occur on the vowel, which is typically the longest segment. Campbell uses standard deviation, and mean duration (both values estimated from the British English corpus SCRIBE), as proxies for segment elasticity, and underlying duration, respectively.

In Campbell's model, compensation effects derive from the dependence of the expansion coefficient (ε) on total elasticity. The more segments within a syllable, the smaller ε, and the less any given segment is expanded. Higher-elasticity segments within the syllable produce the same effect. Conversely, segments with lower elasticity force more lengthening to take place over higher-elasticity segments within the same syllable. Because voiced obstruents have lower elasticity than voiceless, a larger expansion coefficient is required for the syllable closed by the voiced obstruent to reach the same target duration as the syllable closed by the voiceless obstruent. The larger expansion coefficient, in turn, results in a longer vowel. Campbell notes that his model "...appears to account quite simply, *though in fact not completely*, for the lengthening that has been observed in vowels of English before voiced consonants." (Campbell, 1992, p. 218, emphasis ours). An underlying difference in mean duration between voiced and voiceless obstruents comprises part of this voicing effect. However, the relative contribution of this fixed value decreases as syllable length increases. The difference in duration due to the differing elasticities of the two segments, on the other hand, increases as syllable length increases. Expansion is a linear function of σ_T, therefore the difference in expansion, $(\varepsilon_{vd} - \varepsilon_{vl})$, also increases linearly. See Appendix (B) for more details.

Because Campbell does not limit the degree to which voiced obstruents can lengthen, the model under-estimates the voicing effect at the longest durations. Without additional mechanisms, the model also fails to account for the shortest end of the distribution, predicting, in fact, that the voicing effect should reverse under compression. Both of these mismatches are due to the use of a linear expansion function in regions of the duration space that do not behave linearly.

3.2 Competing Timing Constraints

Although Campbell (1992) does not explicitly require all syllables to be the same length, the fact that target syllable durations are strictly enforced means that any two syllables can be set to the same duration. In which case, the difference in vowel duration must be equivalent to the difference in obstruent duration between any two VC_1, VC_2 minimal pairs. Campbell additionally predicts that vowels in open syllables will be longer than vowels in closed syllables, and vowels in closed syllables with complex codas will be shorter than vowels in syllables with simplex codas, in both cases, by exactly the duration of the coda consonant. 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. However, while isochronic tendencies exist, it has become clear that uniform timing for syllables is not consistently enforced in English, or in any other language that has been investigated (see Krivokapić (pted) for a review).

One possible reason for imperfect compensation is that isochrony operates at a higher prosodic level than the one being measured. Port et al. (1987) find that moras in Japanese, when produced in isolation, can vary quite widely in duration, yet the words in which those moras appear are much more uniform in length. Small timing adjustments appear to be made at numerous locations within the word, and not necessarily at mora boundaries. Something similar might be true in English, where syllables produced in isolation are clearly not all of the same duration (cf. Chen 1970). On the other hand, if isochrony itself is driven by a pressure for a uniform rate of information transfer, then it could be the case that true isochrony only holds over semantically defined units, such as phrases, or entire utterances (see, e.g., Aylett and Turk 2004, Levy and Jaeger 2007).

Nevertheless, there are temporal trade-offs observable at the syllable level in English that can be modeled without assuming isochrony at any level. This is appealing for "compensatory" phenomena that range widely in their degree, from extreme under-compensation, to significant over-compensation (e.g., Elert 1965, Kristoffersen 2000, Kavitskaya 2002, Munhall et al. 1992, Kim and Cole 2005). It is a central premise of Articulatory Phonology (AP) that inter-gestural and inter-segmental timing relations are the product of the interaction of competing articulatory pressures, none of which need be directly compensatory in nature (e.g., Browman and Goldstein). Phrase-final lengthening, polysyllabic shortening, and onset-nucleus length trade-offs, have been successfully modeled in this framework[5] (Browman and Goldstein 1988, Nam and Saltzman 2003, Saltzman et al. 2008, O'Dell and Nieminen 1999). The hypothesis that we will adopt is that apparent voicing compensation can be treated in an analogous way: as the optimal solution to a set of conflicting timing constraints.[6]We will also argue that prominence-based compensation effects derive from the same source: differences in segmental elasticity.

3.2.1 Compensation for Number of Elements

One of the central results of AP is the so-called c-center effect, which explains apparent vowel duration differences between syllables with simplex versus complex onsets (Browman and Goldstein 1988, Nam and Saltzman 2003, Saltzman et al. 2008). This apparent compensation is less than total, with syllable duration increasing with each additional consonant in the onset. That

[5] Coupled-oscillator systems represent a type of constraint conflict; each individual oscillator has a different preferred oscillation frequency, and it is not possible to satisfy both frequency preferences (constraints) in a joint system. Instead, a "compromise" frequency for the system is adopted that is somewhere between the two individual frequencies, violating each constraint minimally (according to their relative weights).

[6] 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.

compensation is only partial is explained by the fact that such differences are not the result of actual compensation. Syllable organization is a function of preferred phasing relationships between successive articulatory gestures. In an English CV syllable, the vocalic gesture is initiated at the midpoint of the consonant gesture. However, a conflict arises when multiple consonants share the same preferred phasing with respect to the following vowel. Satisfying all of them would result in complete merger, or masking of the consonantal gestures. At the same time, the consonants have different timing preferences with respect to one another. The result is essentially a compromise in which the timing for each consonant is shifted earlier or later by an amount that allows for both the C-V and the C-C phasing to deviate minimally from their preferred values, while preserving sufficient acoustic cues for all segments. A shift earlier for the first consonant has the effect of lengthening the syllable somewhat, while a shift later for subsequent consonants has the effect of masking more of the vowel. In the latter case, the vowel is acoustically shorter, but not articulatorily.[7]

Apparent compensation is found at a number of different levels: words are shorter, the more words there are in the same utterance, stems are shorter the more affixes are attached (Lehiste 1972),[8] and stressed syllables are shorter, the greater the number of following unstressed syllables (e.g., Fowler 1981). As with segment-level compensation, stressed syllable duration consistently under-compensates, such that total duration increases (non-linearly) for each additional unstressed syllable (Lindblom and Rapp, 1971, Kim and Cole, 2005). So-called

[7] In the case of coda clusters, it has been proposed that something similar to a c-center effect could account for the compensatory behavior (Fowler et al. 1986, Munhall et al. 1992). However, this seems to contradict an earlier finding that only the initial consonant of a coda cluster is coordinated with the vowel, while the remainder are only coordinated with their immediately preceding consonant (e.g., Browman and Goldstein 1988). The addition of more consonants should thus make the syllable longer but have little to no impact on the acoustic duration of the vowel. Nevertheless, articulatory overlap, leading to acoustic masking, may help explain the duration difference between an open versus a closed syllable. This mechanism alone, however, would not be able to account for the wildly varying degree of compensation (anywhere from 13 to 100 ms) reported by Maddieson (1985).

[8] Lehiste (1972) finds that stems are shorter in the affixed form than in isolation (e.g., sleep/sleepy). Furthermore, "shortening" increases with the addition of a second affix. This effect interacts with final voicing, such that the amount of "shortening" for voiced-final stems is greater (both absolutely, and proportionally) than that for voiceless-final. A difference is also found between inherently longer and shorter vowels, with longer being more "compressible". Words are also shorter, the more words in a given utterance, and this interacts with position, with words successively longer the closer they are to the end of the utterance (and thus the phrase boundary).

polysyllabic shortening has also been modeled as the result of competing constraints, instantiated as a coupled oscillator system in which the preferred frequency of the oscillator at the lower level of the prosodic hierarchy (e.g., the syllable level) conflicts with the preferred frequency of the oscillator at the higher level of the prosodic hierarchy (e.g., the foot), resulting in a frequency intermediate between the two (O'Dell and Nieminen 1999).

3.2.2 Compensation for Intrinsic Duration

Vowels, as the more expandable segments, seem to compensate for consonant duration, but consonants rarely, if ever, seem to compensate for inherent vowel duration differences.[9] Without compensation, syllable duration varies significantly by vowel type: syllables with low vowels are generally longer than those with high vowels; syllables with tense vowels tend to be longer than syllables with lax vowels; stressed syllables are longer than unstressed syllables (Peterson and Lehiste, 1960, Sharf, 1962, De Jong, 2004). However, it appears that compensation can occur between two syllables of inherently different duration within the same word.

So-called "prominence-based compensation" typically involves a length trade-off between a stressed and an unstressed vowel. 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, Dutch, the penultimate syllable of the final word also sometimes experiences significant lengthening. This happens only when the final syllable is unstressed, or contains a schwa vowel (see also, Turk and Shattuck-Hufnagel, 2007). Katsika (2016) reports a similar finding for Greek, with the articulatory mechanisms for final lengthening appearing to shift towards a stressed penultimate syllable.[10]

[9] Munhall et al. (1992) find a very small difference (on the order of a few milliseconds in consonant duration following vowels of different lengths

[10] Within the Articulatory Phonology framework, prominence-based compensation would arise from the interaction between two different types of basic gestures: the μ-gesture, that is associated with stressed syllables, and the π-gesture that is associated with boundary edges. Both are conceptualized as localized "clock-slowing" gestures that result in lengthening (e.g., Byrd and Saltzman 2003, Saltzman et al. 2008). Thus the interaction, or coupling, between

The voicing effect can be described in similar terms: lengthening (often due to a phrase-final boundary) “shifts” to earlier segments (the vowel) when the final segment (the voiced obstruent) cannot be lengthened sufficiently.[11] That unstressed vowels and voiced obstruents should be “compensated” for seems to be a consequence of their relatively shorter durations. However, as we have seen, voiced obstruents are not consistently shorter than voiceless obstruents. The relationship between the two distributions, furthermore, may be one that is typical of real speech: little to no difference between segments at the majority of durations, with inherently longer segments differentiating themselves only in the upper tail of the distribution (see also Crystal and House, 1988, Campbell, 1992). The duration distributions for the class of tense non-high vowels, tense high, and lax, are plotted in Figure 5, as an example.[12] Such a relationship would explain why “compensations” for short segments and short syllables is observed primarily in phrase-final position.

Under our proposal, it is the elasticity of each segment that is specified. Apparent compensation is modeled as the result of a conflict between a target duration at the syllable or word level, and the duration preferences of the individual segments. Duration differences are not *preserved* under lengthening (e.g., Peterson and Lehiste, 1960, Sharf, 1962, De Jong, 2004). Rather, they emerge, and increase, under lengthening, disappearing as durations get shorter.

3.3 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. The results reported here are for VC syllables. See Appendix (C) for the treatment of CV syllables, and the “voicing” effect in onset position.

Constraints on segment duration are implemented as Normal probability distributions with

these two gesture types should, in principle, result in prominence-based compensation. However, as far as we are aware, this has not yet been explicitly modeled.

[11] 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.

[12] The data come from the set of obstruent-final CVC words used in the voicing analysis.

a characteristic mean and variance. These distributions act as a type of stochastic constraint in that they assign the highest probability to their preferred duration value, but non-zero probabilities to other, less-preferred durations. In a competition with conflicting constraints on duration preferences, a longer or shorter value will be selected, depending on the full set of constraints and their weights. Weights are reflected in the size of the variance. This is not a measure of the actual duration variance (since that is determined by the interaction of all constraints), but governs how quickly probability decreases away from the mean. All else being equal, a segment with a broader distribution will be lengthened or shortened more than a segment with a narrower distribution. Thus variance also acts as a measure of elasticity. Unlike in Campbell's model, however, the effective resistance to lengthening is not constant. The further one gets from the mean, the lower the probability becomes. This means that expansion and compression are non-linear.

The three segment-level constraints are shown graphically in Fig. 6. Voiced and voiceless obstruent distributions have the same mean value in these simulations, differing only in their variance. The competition in this model is realized through maximization of the joint probability function over all constraints. This function exhibits the desired behavior: the output is only optimal if a decrease in probability for any given variable is accompanied by a greater increase in probability for one or more other variables. Target syllable duration is treated as a random variable, and a constraint for matching syllable duration competes with those for matching preferred vowel and consonant duration.

In addition, two inter-segment timing 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 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. The model searches for the durations of the coda consonant (D or T) and vowel (V) that result in the highest joint probability over this set of constraints.[13] The search

[13] Following Browman and Goldstein (1988) inter alia, we assume that there is a preferred timing relationship for a

is conducted in a brute force manner, by simply trying all possible combinations of values. However, the search space is restricted within a certain range, and a fixed step size is used. The result is an approximation to the true maximum, within the resolution of the step size. Each variable is assumed to be independent, so the joint probability is given as the product of the individual probabilities. See Appendix (C) for further details of the model.

Figure 7 shows the behavior of the duration variables over a representative range of target syllable durations. Voiced and voiceless obstruents (black and red solid lines, respectively) are more or less identical in duration for a sizeable range of target syllable durations; preceding vowel durations (black and red 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 optimal duration for the voiced obstruent falls below that of the voiceless obstruent at longer durations. At the same time, the pre-voiced vowel duration starts to diverge from the pre-voiceless. Because of the pressure to match the target syllable duration, a somewhat longer vowel duration, preceding a somewhat shorter voiced obstruent, maximizes the joint probability. Like the consonant duration difference, the vowel duration difference will continue to increase, meaning that the magnitude of the voicing effect will increase with increasing duration.

Because the constraint to match target duration competes with other constraints, a given syllable does not always match the target exactly. And because coda elasticity affects the outcome, voiced and voiceless syllables at the same target syllable duration do not necessarily have the same actual syllable durations. Thus this model allows for under- and over-compensation. This occurs only when imperfect matching would increase over-all probability. The weight of the constraint that enforces matching can be modified by changing the variance of the associated probability distribution. A lower variance maps to a higher weight. As the variance approaches zero, compensation becomes total.

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.

For intermediate variance values, the following general behavior can be seen. In the expansion regime, 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 can be seen in Fig. 7 in the difference between the gray line ($\sigma = \sigma_T$), and the two dotted lines that indicate actual syllable duration. The decrease in compensation 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. 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, 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 from a Normal distribution of target syllable durations, durations that fall mostly in the range where there is a negligible difference in consonant duration. This sample is represented by the light blue vertical bars in Fig. 7.[14] The vowel and consonant duration distributions resulting from this sample are shown in Figure 8.

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. A significant voicing effect only emerges at longer absolute durations. The magnitude of the effect increases with increasing duration. The difference in vowel duration is directly reflected in the difference between obstruent durations. These results are achieved without any directly compensatory mechanism. The model can also capture the interaction between the voicing effect and vowel quality, using lower elasticity parameters for inherently shorter vowels. See Appendix (C).

[14] The minimum duration of a nucleus was set at 30 ms. All durations less than 30 ms were set to 0.

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. Similarly, pre-pausal lengthening can also produce a significant voicing effect. A particularly large final lengthening effect in English (e.g., Delattre 1966), we hypothesize, may be largely responsible for the particularly large voicing effect in this language.

4 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,[15] 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 that demonstrate diminished lengthening for voiced obstruents relative to voiceless obstruents as speaking rate decreases, and a difference in preceding vowel duration that mirrors the difference between the obstruent durations.

4.1 Procedure

All participants were undergraduate students at The Ohio State University who were given course credit for completing the experiment. 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.[16] 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 bock, participants

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

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

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, 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.[17]

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 varied across vowel quality (o or i),

[17] 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.

consonant manner (stop or fricative), and final consonant place (coronal or labial): feet/feed, thief/thieve, lobe/lope, and doze/dose.

4.2 Data Selection and Annotation

Each participant produced approximately 8 tokens of each word at each rate. Of those 8, 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 repetition rate, it was possible to segment both the closure and the release interval for each stop. However, at the fastest rates, final stops did not always 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 first dropped to background levels. 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 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.

Occasionally the voiced stops and fricatives at the slower repetition rates were produced with a final epenthetic schwa. There were 27 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. For two participants, they were practically identical in the production of final s and z. All dose/doze tokens for those two participants were removed. Four 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. This left 23 participants, and 875 total tokens.

The data for the first two word pairs (feet/feed, thief/thieve) were distributed among three undergraduate research assistants. 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 two different RAs. Measurement verification was conducted in the same way. Praat (Boersma and Weenink 2009) was used for segmentation and annotation.

4.3 Results

In Fig. 9 final stop durations 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, VOT, and the sum of the two (TDur). 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.[18]

Vowel duration (red), total obstruent duration (green), and total rhyme duration (vowel + TDur) are shown in Figure 10 for the full set of words. Visual inspection shows that larger vowel durations were reached by the voiced member of each minimal pair (bottom panels), while larger obstruent durations were reached by the voiceless member (top panels). There is also a larger difference between consonant and vowel durations for voiced-final tokens across all repetition

[18] Whether total stop duration, closure duration, or aspiration duration is the relevant durational measure may well be language-specific. Durvasula and Luo (2012) report longer vowel durations preceding voiced stops, as well as longer vowel durations preceding aspirated stops in Hindi (which has a four-way contrast system). Both effects can be attributed to expandability if the relevant duration measure is closure duration. Closure durations for aspirated stops were found to be shorter than unaspirated, and closure durations were also shorter for voiced than voiceless stops. Total stop durations were not reported. If aspiration duration is significantly shorter than closure duration and/or aspiration does not participate in lengthening, then these results are consistent with our hypothesis. For the 3-way voiced, voiceless, and ejective contrast in Georgian, Beguš (2017) finds that closure duration and VOT both negatively correlate with preceding vowel duration, with ejective stops intermediate between voiced and voiceless in terms of both consonant duration and preceding vowel duration.

rates, and that difference increases with decreasing repetition rate.

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. 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 3. Only significant effects are shown. Adding voicing to this model did not improve fit.

A separate model of vowel duration as a function of rate and voicing behaves very similarly. 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. However, adding consonant duration to this model *does* significantly improve model fit. In fact, adding consonant duration renders voicing no longer significant, except in interaction with rate. The significant interaction of rate and voicing can be explained by the fact that pre-voiced vowels are consistently longer than pre-voiceless only at slower rates. See Table 4. Only significant effects (other than voicing) are shown.

If consonant duration is a better predictor of vowel duration than voicing, then any positive effect of voicing on vowel duration is accounted for by the negative effect of voicing on consonant duration. Furthermore, the fact that consonant duration is a better predictor of vowel duration at slower rates derives from the fact that consonant duration across voiced and voiceless pairs diverges more the longer the word gets. The model of consonant duration as a function of voicing confirms this analysis. See Table 5. 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 (stops are the reference level). 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. Two interactions between manner and rate also reached significance: a positive interaction with linear rate; and a negative interaction with quadratic rate. We interpret these two effects to mean that the difference in duration between stops and fricatives is a nonlinear function of rate. Only significant interactions are shown.

These trends are highlighted in Figure 11, where the difference in duration between voiceless and voiced members of a given minimal pair is plotted by rate (in order to do a paired analysis, only words with all 5 rates, for both voiceless and voiced members, were plotted; a total of 820 tokens). There is a clear negative correlation between the *difference* in duration of voiceless and voiced consonants, and the *difference* in duration of their preceding vowels. Vowel duration difference (voiceless-voiced) is also negatively correlated with repetition rate: the voiceless-voiced difference becomes more negative as speaking rate decreases.

In a 3-factor model of vowel difference ($V_{VL} - V_{VD}$) with no interaction terms, rate, consonant difference ($C_{VL} - C_{VD}$), and manner are all significant. In a fully crossed model, the correlation between rate and consonant duration difference can be clearly seen. In this model there are significant main effects only of $C_{VL} - C_{VD}$, (a larger positive difference contributes to a larger negative vowel duration difference) and manner (on average, words closed by fricatives had voicing effects of 98 ms, versus 52 ms for words closed by stops). Rate fails to reach significance. However, a significant three way interaction indicates that rate does significantly contribute to model fit. Thus the voicing effect is shown to be larger for fricatives than stops, and for larger negative values of $C_{VL} - C_{VD}$, both of which are enhanced at the slowest speeds. See Table 6. Only significant factors are shown.

In a fully crossed mixed-effects model for rhyme duration, as a function of rate, manner and voicing, only rate (linear) and manner were significant. Rhyme durations for fricatives were, on average, 71 ms longer than for rhymes containing stops. Rhyme duration *differences* were larger for fricatives than stops as well. Voiced fricative rhymes were 25.5 ms longer on average than voiceless fricative rhymes (a significant difference under a two-sided t-test: t = -4.1781, df = 208, p-value = 4.326e-05), whereas voiced stop rhymes were only 2.8 ms longer than voiceless stop rhymes.

The full set of 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,[19] 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 highly 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 significant obstruent duration differences only emerge at slow rates (long durations), and vowel duration differences derive from those differences, rather than depending on phonetic voicing, or an abstract phonological voicing feature.

It should be noted that the experimental task is highly unnatural, and most likely biases more towards uniform syllable duration than natural speech contexts. Thus these results probably

[19] 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.

over-estimate the degree to which vowel and consonant duration are traded off. Nevertheless, compensation was still not total, as significant differences in rhyme duration were found.

5 Further Tests of The Expandability Hypothesis

In the previous sections we have shown that vowel duration depends more on coda duration than on 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 gestural organization through competition can qualitatively capture the duration trade-offs between consonant and vowel in production. In this section we test the Expandability Hypothesis against the perception literature and present a number of testable predictions generated by our hypothesis.

5.1 Perception of voicing in final position

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. However, we will show that not only is this 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. It is primarily for stops that preceding vowel duration has been characterized as a contrastive cue. This is plausible because there are often very limited cues to stops in final position.

While listeners can, and do, make use of preceding vowel duration to identify ambiguous following stops, they make use of other cues as well. While most studies do not directly test different cues against one another, among those that do, the balance of evidence actually comes down against the effectiveness of the vowel duration cue. Both Raphael (1972) and Crowther and Mann (1992) report that preceding vowel duration is stronger than F1 as a cue to voicing. However, 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 the removal of 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. Raphael (1981) concludes that vowel duration is only a weak cue to voicing for natural stimuli produced in carrier phrases, and that the effectiveness of various cues is strongly context-dependent.

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). However, compared to the number of studies that test perception based on preceding vowel duration alone, there are relatively few that manipulate, or even report, the duration of final stops. These studies, for whatever reason, also tend to be cited less frequently. However, Raphael (1981) found that swapping closure durations for naturally produced "peg" and "peck" effectively switched the voicing percept for the two tokens. Repp and Williams (1985) similarly found an effect of overall closure duration on the perception of voicing on stop-final syllables followed by a stop-initial syllable (e.g, "lab coat" vs. "lap-coat").

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[20]. 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 each other. 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

[20] 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).

predictive.

The compensation model of Section 3 is used to illustrate this hypothesis. See Fig. 12. The duration of the voiceless stop (red 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 (red dashed line – preceding voiceless stop; black dashed line – preceding voiced stop). If the listener is exposed to a relatively short vowel (Fig. 11a: horizontal gray 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 (blue dotted line) that falls close enough to both expected values is assumed be acceptable for either member of the pair, and will not be sufficient to distinguish between the two in the absence of other cues.

For a very long vowel, on the other hand, there is a large difference in the expected durations of the voiced and voiceless stops. See Figure 11b. The same observed stop duration (dotted blue 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. An ambiguous final stop of middling duration becomes less ambiguous as vowel duration increases (speaking rate decreases). The cross-over point occurs when 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. However, the nature of the actual experimental stimuli may bias perception strongly towards the voiceless member of the pair for two reasons. 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. 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.[21]

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.[22] 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 for less extreme durations, and in an explicit comparison task where listeners must label one token as voiced, and one as voiceless. In 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

[21] Although Raphael (1972) tested both “voiced” and “voiceless” final synthetic stimuli, the only difference was that the voiceless lacked any F1 values at all during the transition period. All tokens in both experiments lacked a voicing bar, and contained no release bursts.

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

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 Stops in Initial and Medial Position

Like word-final stops, medial stops are post-vocalic, as well as subject to neutralization of both voicing and aspiration, resulting in productions that are essentially just short periods of silence (oral cavity closure). Vowel duration has also been shown to be a sufficient cue to voicing in medial position. In this literature, however, it is standard to describe the perceptual boundary in terms of the ratio of closure duration to preceding vowel duration (e.g., Lisker 1957a, Port and Dalby 1982, Port 1979, 1981). The C/V ratio also normalizes stop duration relative to estimated speaking rate. This type of normalization presumably also applies in final position, yet because final closure duration is not typically measured, let alone systematically varied, it has become tacitly assumed that C duration plays no role in perception – at least for stops, and pre-pausally. Ultimately, the size of the "voicing" effect seems to be the only real difference between medial and final position. Voiceless closure durations are reported to be from 30-45 ms longer than voiced,[23] and pre-voiced vowel durations, 25-35 ms longer than pre-voiceless (Lisker, 1957b, Sharf, 1962, Davis and Van Summers, 1989). These smaller values are to be expected, given that final lengthening does not apply,[24] and words are, by definition, polysyllabic.

[23] These values are for labial and velar place. Coronals are typically flapped in this environment and show little to no duration differences.

[24] In medial position the consonant's syllabic affiliation is ambiguous. The onset maximization principle (e.g., Clements and Keyser 1983) places a single medial consonant in the onset of the following syllable. However, there is evidence that stress and sonority both affect syllabification, such that a medial consonant will be syllabified in the coda of a preceding syllable if that syllable is stressed, and the following syllable is not (e.g., Treiman et al. 1994;

There is a very large body of work devoted to word-initial stops, in which VOTs in pre-stressed position are typically measured. The relationship between VOT and following vowel duration, however, has been much less studied. When post-stop vowel duration is manipulated in perception experiments, it is almost always done so as part of a speaking rate study in which stop duration is inversely co-varied, making it difficult to determine the relationship between consonant and vowel (e.g., Miller and Baer 1983, Miller and Volaitis 1989, Volaitis and Miller 1992). However, the same resistance to lengthening under decreases in speaking rate in final position is 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 (see Section 3). Viswanathan et al. (2019) report a significant difference in voicing perception only between vowels of 175 and 225 ms., 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: 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.[25] However, the difference in gestural timing relationships between vowel and onset versus vowel and coda is

Eddington et al. 2013). In either case, word-final and phrase-final lengthening do not apply.

[25] 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).

also expected to contribute to this outcome. For both medial and initial stops, a correlation has been found between the voicing category of the stop, and the length of the tautosyllabic vowel. By definition, this is a “voicing” effect, despite the different terms in which these patterns are described. In both cases also, we find that stop duration correlates both with voicing, and, inversely, with vowel duration, making these results broadly consistent with the Expandability Hypothesis. See Appendix (C) for a model of the “voicing” effect in initial position.

5.3 Predictions

If the Expandability Hypothesis is correct, 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. Weismer (1979) and Choi et al. (2016) have actually found that the VOT is longer for voiceless stops word-initially in CVC words when the final stop is voiced, than when it is voiceless.[26] A difference in nasal duration preceding voiced versus voiceless stops has also 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 before a voiced stop.[27] However, no explanation is offered as to why the phasing relationship should be different in the two contexts.

[26] Although we treat onsets as external to timing considerations, this is clearly a simplification. Onsets do contribute something to syllable duration, even if they play a smaller role in prosodic phenomena than the coda.

[27] Pycha and Dahan (2016) find a difference in the ratio of nucleus to offglide for the diphthong a͡ɪ before voiced versus voiceless stops that they attribute to the same phenomenon. They also show in an eye-tracking study that listeners can use the nucleus duration in some cases to predict the voicing of the coda stop.

This difference, however, can be accounted for under the Expandability Hypothesis if expansion compensation at the word (or syllable) level acts to both lengthen individual gestures, as well as separate them, as has been found under decreases 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, results in less vowel nasalization. Thus, a highly nasalized vowel is more likely to occur preceding a [t] than a [d].

Our model also predicts that a change in the perception of voicing should lead to a change in the perception of speaking rate. During the course of the vowel production, it is assumed that a hypothesis about both speaking rate and following segment duration is generated by the listener. For a vowel that is so long that it creates a strong expectation for a following voiced stop, a certain speaking rate is also inferred (represented by the x-intercept of the vertical line on the left in Figure 11b). If listeners subsequently experience an unambiguously voiceless stop (e.g., strongly aspirated), a simultaneous change in their experience of the speaking rate of the entire syllable should occur (shifting to the x-intercept of the right-hand vertical line in Figure 11b). The voiceless stop should indicate that the speaking rate is actually slower than previously supposed, not (or not only) because the stop itself adds duration to the syllable, but because a vowel of the given duration, preceding a voiced stop, is expected to occur at a faster speaking rate than a vowel of the same duration, preceding a voiceless stop. To the best of our knowledge, there are no data yet that directly test this prediction.

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. Sharf (1964) explicitly found that vowel duration differences were approximately the same whether words were produced normally or whispered, thus demonstrating that the effect was independent of actual vocal fold vibration.[28]

[28] This result is usually cited as evidence for the "linguistically determined" (i.e., phonologized) nature of the vowel

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. In principle, any apparent temporal compensation phenomenon could potentially be modeled using the competing constraints framework (see Section 3.2). For example, characteristic differences in stop duration across place of articulation are fairly robust, and have also been associated with following vowel duration differences. However, the differences are very small, on the order of 10 ms for following vowels and 5 ms for the consonants themselves (Fant 1973, House 1961, Luce and Charles-Luce 1985, Elert 1965). Vowel duration differences preceding stops at various places of articulation, are similarly small, on the order of 5-15 ms. (Elert 1965, Peterson and Lehiste 1960). It is not clear whether these differences in consonant duration should be attributed to differences in characteristic duration, or difference in elasticity, or both.

All else being equal, we might 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 (Peterson and Lehiste, 1960, Umeda, 1975), which is the opposite of what one would expect for a sonorous segment like a nasal. The phasing relationship between vowel and coda, however, may be quite different in the case where the two gestures can overlap significantly without masking. Thus vowel durations may appear longer before nasals because there is significantly more coarticulation than occurs with other consonants (the length of the nasals themselves was not reported in these studies). If this is correct, then the vowel should be acoustically highly nasalized.

It has also been consistently found that vowels preceding voiced fricatives are the longest, while vowels preceding voiceless fricatives are somewhat longer than those preceding voiceless stops (Umeda, 1975, Peterson and Lehiste, 1960).[29] Furthermore, the voicing effect has been

duration cue.

[29] 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

reported to be larger for fricatives than for stops (e.g., House 1961). Although our production experiment was not designed to explicitly test fricatives against stops, the results are in line with these findings. Vowel durations were longest before voiced fricatives, and a larger voicing effect was found for fricatives than stops (98 ms difference in vowel duration, versus 52 ms). Vowels were also longer before voiceless fricatives than voiceless stops. However, the Expandability Hypothesis predicts that preceding vowel durations should be similar for voiced stops and fricatives, given that voiced fricatives were only 9 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 33 ms longer than voiceless stops.

These results are driven in large part by the dose/doze pair, and may be partially explained by the fact that the short (d) onset (as compared to l, θ and f onsets in the other words) leads to longer vowel durations at all rates. A longer baseline vowel duration may then be compounded over increasingly slower rates. Nevertheless, the results trend in the same direction for the thief/thieve pair, and are reported in other work. Therefore, the differences between the behavior of fricatives and stops remains to be adequately explained by any hypothesis. This will minimally require follow-up work using carefully balanced stimuli.

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, several speculative explanations in this section rely on assumptions 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

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.

occur without apparent compensation on vowels that are inherently short (see Appendix C). There are also predictions about differences across final, medial, and initial position that cannot be sufficiently 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 most 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. From the vast literature, only a handful of studies are regularly cited, and they tend to be those that demonstrate either strong categorical perception (e.g., Raphael 1972) or large vowel duration differences in production (e.g., Mack 1982). 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). In this paper we have synthesized the larger literature, demonstrating that vowel length differences correlated with final obstruent voicing are dependent on a number of factors that interact in a more complex way.

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). We confirmed both these results using the Buckeye Speech Corpus, finding no over-all effect of final-obstruent voicing on vowel duration (in the predicted direction), but a significant negative effect of obstruent duration. We hypothesized that the lack of a voicing effect was due to the fact that voiced and voiceless obstruents themselves were not consistently different in duration in many cases. The data suggested that significant duration differences across voicing class only emerge at the very high end of the duration distribution, when analysis is restricted to low frequency words containing low tense vowels, produced at a slower speaking rate, and/or in a pre-pausal context.

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,[30] demonstrating that the difference in vowel duration increased in step with the inverse duration difference for obstruents. 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 when it was significantly correlated with duration (i.e., at the slower rates).

Aspiration, and actual voicing, have 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). Additionally, obstruent duration itself has been shown to affect perception in final position (Denes 1955, Raphael 1981, Repp and Williams 1985), just as it does in 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 is a cue to obstruent duration itself, and only predictive of voicing class within a certain upper range of durations. We have offered a proposal that accommodates this full set of results, as well as additional related 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.

(3) The Expandability Hypothesis

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

[30] 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.

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. The 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, pted).

Our model provides a proof of concept for deriving the voicing effect from a set of general-purpose timing constraints. Our account also covers much more empirical ground than explanations of the voicing effect that are based on actual vocal fold vibration, or articulatory effort. We are able to unify the treatment of the contrast across word and syllable position, and draw connections between effects based on differences between consonant elasticity, and those based on differences between vowel elasticity. Our explanation for the voicing effect also also has ramifications for theories of contrastive features.

6.1 The Right "Voicing" Features

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 to account for the surface realizations of the contrast it is necessary to treat the phonological voicing feature as distinct from the phonetic feature of the same name. The first must be only an abstract label which is then transformed through a series of rules to the actual phonetic realization of the sounds. For example, in absolute initial position the $/-voice/$ stop is realized phonetically as voiceless and aspirated, while the $/+voice/$ member may be realized as voiceless unaspirated. In intervocalic pre-stress position $/-voice/$ is also realized as aspirated, but the $/+voice/$ segment is likely to be phonetically voiced in this environment. In intervocalic post-stress position, both phonemes may be voiced, and/or flapped. Following */s/* pre-vocalically, both may be voiceless and unaspirated. In syllable-final position both stops may be realized as voiceless and unreleased. And, of course, preceding vowel duration is greater for underlyingly voiced obstruents.

Whatever other features they possess, we have argued that segmental elasticity must comprise a critical part of the specification of the voiced/voiceless distinction in English. Rather than a discrete allophonic rule, apparent vowel lengthening, we have argued, is derived from elasticity differences between the two types of obstruent, and varies considerably as a function of speaking rate, sentence and word position, stress, and other factors. Listeners are able to exploit the correlation between preceding vowel duration and obstruent voicing in order to discriminate the contrast under certain conditions. However, this is not particularly noteworthy, given that the number of acoustic cues to the contrast 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 vowel duration is frequently characterized as a phonological "voicing" feature, but not closure duration or F0, 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 Contrast and Allophony

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. Partial, or total, devoicing is also a possible outcome. That "voiced" stops in English are now frequently devoiced means that the observed duration differences are not the direct result of physiological constraints. Furthermore, we have shown that duration differences between voiced and voiceless are not consistently present in normal speech. The inherent difference between the two seems to reside, not in duration per se, but in the degree and rate of expansion. This conclusion is based on evidence that voiced obstruents show an apparent resistance to lengthening. In representational terms, our account requires voiced obstruents (and, indeed, all segments) to have a specification of something equivalent to elasticity. 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, either at the phonological or phonetic level.

The chain of reasoning just described, however, turns out to directly conflict with the dominant view of the role of prominence in phonological theory. While the phonetic realization of underlyingly contrastive features will differ by context, the most prominent environment, usually initial pre-stress position, is assumed to most faithfully reflect those features. 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. 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.[31] The phonetically voiceless stops in French and Thai fail to lengthen significantly with decreased speaking rate, and are therefore unspecified for laryngeal features, while the phonetically short lag/voiced stops in English are the unspecified member of the contrast[32] (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 (at either the phonetic or phonological level). The following set of results has been taken as evidence that preceding vowel duration is purposefully manipulated by speakers to enhance a laryngeal contrast: that the effect of stress is smaller for pre-voiceless than pre-voiced vowels in English De Jong (2004); that short pre-voiced vowels lengthen less that long pre-voiced vowels in order to 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 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).

[31] According to Beckman et al. (2011), Swedish exhibits an unusual 2-way contrast in which both members of the contrast are specified: one for [voice], and the other for [spread glottis].

[32] English is usually described as a [spread glottis]/Ø system

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 expandability; 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.[33]

The assumption that prominent contexts are somehow privileged is also at the heart of basic analytical decisions about which features are underlying, and which derived. Contrastive features are assumed to be most fully realized in word-initial, pre-stressed environments, when they are spoken slowly, or adjacent to a high-level prosodic boundary. This view, however, requires potentially extensive transformations of such underlying forms to the more frequent, non-prominent contexts of normal speech. If we reverse this relation, however, then very slow hyper-articulated speech is the exception, augmenting segments with intense aspiration and especially long durations that are not typical of the contrast in general. Large differences in preceding vowel duration are, almost exclusively, the product of this kind of 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.

[33] Shortening effects, such as those reported by Lehiste (1972), can be explained by the fact that a larger proportion of the shortening occurs over the voiceless than the voiced stop, resulting in a larger reduction of the pre-voiced vowel.

7 References

Abdelli-Beruh, N. (2004). The stop voicing contrast in french sentences: Contextual sensitivity of vowel duration, closure duration, voice onset time, stop release and closure voicing. *Phonetica*, 61(4):201–219.

Ainsworth, W. A. (1972). Duration as a cue in the recognition of synthetic vowels. *The Journal of the Acoustical Society of America*, 51(2B):648–651.

Allen, G. D. (1975). Speech rhythm: its relation to performance universals and articulatory timing. *Journal of phonetics*, 3(2):75–86.

Allen, J. S. and Miller, J. L. (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(4):2031–2039.

Aylett, M. and Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1):31–56.

Bailey, P. J. and Summerfield, Q. (1980). Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, 6(3):536 – 563.

Beckman, J., Helgason, P., McMurray, B., and Ringen, C. (2011). Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics*, 39(1):39–49.

Beckman, J., Jessen, M., and Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics*, 49(2):259–284.

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, 85(4):785–821.

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., and Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America*, 133(4):2350–2366.

Beguš, G. (2017). Effects of ejective stops on preceding vowel duration. *The Journal of the Acoustical Society of America*, 142(4):2168–2184.

Belasco, S. (1958). Variations in vowel duration: Phonemically or phonetically conditioned? *The Journal of the Acoustical Society of America*, 30(11):1049–1050.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1):92–111.

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

Boersma, P. and Weenink, D. (2009). Praat: Doing phonetics by computer (Version 6.0.36). computer program.

Braunschweiler, N. (1997). Integrated cues of voicing and vowel length in German: A production study. *Language and Speech*, 40(4):353–376.

Browman, C. P. and Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, 3:219–252.

Browman, C. P. and Goldstein, L. M. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, 45(2–4):140–155.

Browman, C. P. and Goldstein, L. M. (1990). Tiers in articulatory phonology, with some implications for casual speech. In Kingston, J. and Beckman, M. E., editors, *Papers in Laboratory Phonology I: Between the grammar and the physics of speech*, pages 341–376. Cambridge University Press, Cambridge.

Byrd, D., Kaun, A. R., Narayanan, S., and Saltzman, E. (2000). Phrasal signatures in articulation. In Broe, M. B. and Pierrehumbert, J. B., editors, *Papers in Laboratory Phonology V*, pages 70–87. Cambridge University Press, Cambridge.

Byrd, D., Lee, S., Riggs, D., and Adams, J. (2005). Interacting effects of syllable and phrase position on consonant articulation. *The Journal of the Acoustical Society of America*, 118(6):3860–3873.

Byrd, D. and Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, 26:173–199.

Byrd, D. and Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31(2):149–180.

Cambier-Langeveld, G. M. (1997). The domain of final lengthening in the production of Dutch. *Linguistics in the Netherlands*, 14(1):13–24.

Cambier-Langeveld, G. M. (2000). *Temporal marking of accents and boundaries*. Den Haag: Holland Academic Graphics.

Campbell, W. N. (1990). Timing invariance in read speech. In *Speaker Characterization in Speech Technology*, pages 78–83.

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

Campbell, W. N. and Isard, S. D. (1991). Segment durations in a syllable frame. *Journal of Phonetics*, 19(1):37–47.

Catford, J. C. (1977). *Fundamental problems in phonetics*. Midland Books.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22(3):129–159.

Cho, T. (2016). Prosodic boundary strengthening in the phonetics–prosody interface. *Language and Linguistics Compass*, 10(3):120–141.

Cho, T. and Jun, S.-A. (2000). Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean. *UCLA Working Papers in Phonetics*, pages 57–70.

Cho, T. and Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, 37(4):466–485.

Choi, J., Kim, S., and Cho, T. (2016). Phonetic encoding of coda voicing contrast under different focus conditions in l1 vs. l2 english. *Frontiers in psychology*, 7.

Clements, G. N. and Keyser, S. J. (1983). *CV Phonology: A generative theory of the syllable*. MIT Press.

Coretta, S. (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(1).

Crowther, C. S. and Mann, V. (1992). Native language factors affecting use of vocalic cues to final consonant voicing in English. *The Journal of the Acoustical Society of America*, 92(2):711–722.

Crystal, T. H. and House, A. S. (1982). Segmental durations in connected speech signals: Preliminary results. *The Journal of the Acoustical Society of America*, 72(3):705–716.

Crystal, T. H. and House, A. S. (1988). Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America*, 83(4):1553–1573.

Cuartero Torres, N. (2002). *Voicing assimilation in Catalan and English.* PhD thesis, Universitat Autonoma de Barcelona.

Davis, S. and Van Summers, W. (1989). Vowel length and closure duration in word-medial VC sequences. *Journal of Phonetics*, 17(4):339–353.

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

De Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America*, 97(1):491–504.

De Jong, K. J. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics*, 32(4):493–516.

De Jong, K. J. and Zawaydeh, B. (2002). Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics*, 30(1):53–75.

Delattre, P. (1962). Some factors of vowel duration and their cross-linguistic validity. *The Journal of the Acoustical Society of America*, 34(8):1141–1143.

Delattre, P. (1966). A comparison of syllable length conditioning among languages. *IRAL-International Review of Applied Linguistics in Language Teaching*, 4(1-4):183–198.

Denes, P. (1955). Effect of duration on the perception of voicing. *The Journal of the Acoustical Society of America*, 27(4):761–764.

Derr, M. A. and Massaro, D. W. (1980). The contribution of vowel duration, F0 contour, and frication duration as cues to the /juz/-/jus/ distinction. *Perception & Psychophysics*, 27(1):51–59.

Durvasula, K. and Luo, Q. (2012). Voicing, aspiration, and vowel duration in Hindi. In *Proceedings of Meetings on Acoustics 164ASA*, volume 18, page 060009.

Eddington, D., Treiman, R., and Elzinga, D. (2013). Syllabification of American English: Evidence from a large-scale experiment. Part i. *Journal of Quantitative Linguistics*, 20(1):45–67.

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

Fant, G. (1973). Stops in cv-syllables. Technical report, Dept. of Speech, Music and Hearing: Quarterly Progress and Status Report.

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

Fischer, R. M. and Ohde, R. N. (1990). Spectral and duration properties of front vowels as cues to final stop-consonant voicing. *The Journal of the Acoustical Society of America*, 88(3):1250–1259.

Fitch, H. L. (1981). Distinguishing temporal information for speaking rate from temporal information for intervocalic stop consonant voicing. Technical report, Haskins Laboratory.

Flege, J. (1979). *Phonetic interference in second language acquisition*. PhD thesis, Indiana University.

Fosler-Lussier, E. and Morgan, N. (1999). Effects of speaking rate and word frequency on pronunciations in conversational speech. *Speech Communication*, 29(2-4):137–158.

Fougeron, C. and Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America*, 101(6):3728–3740.

Fowler, C., Munhall, K., Saltzman, E., and Hawkins, S. (1986). Acoustic and articulatory evidence for consonant-vowel interactions. *The Journal of the Acoustical Society of America*, 80(S1):S96–S96.

Fowler, C. A. (1981). A relationship between coarticulation and compensatory shortening. *Phonetica*, 38(1-3):35–50.

Fox, R. A. and Terbeek, D. (1977). Dental flaps, vowel duration and rule ordering in American English. *Journal of Phonetics*, 5(1):27–34.

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

Gahl, S., Yao, Y., and Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66(4):789–806.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *The Journal of the Acoustical Society of America*, 63(1):223–230.

Gimson, A. C. (1970). *An introduction to the pronunciation of English*. Hodder Arnold, London.

Halle, M. and Stevens, K. (1967). Mechanism of glottal vibration for vowels and consonants. *The Journal of the Acoustical Society of America*, 41(6):1613–1613.

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

Harris, M. and Umeda, N. (1974). Effect of speaking mode on temporal factors in speech: Vowel duration. *The Journal of the Acoustical Society of America*, 56(3):1016–1018.

Hillenbrand, J. M., Clark, M. J., and Houde, R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, 108(6):3013–3022.

Hillenbrand, J. M., Ingrisano, D. R., Smith, B. L., and Flege, J. E. (1984). Perception of the voiced–voiceless contrast in syllable-final stops. *The Journal of the Acoustical Society of America*, 76(1):18–26.

Hofhuis, E., Gussenhoven, C., and Rietveld, T. (1995). Final lengthening at prosodic boundaries in Dutch. volume 1, pages 154–157. Stockholm: Stockholm University.

Hogan, J. T. and Rozsypal, A. J. (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(5):1764–1771.

Hooper, J. B. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In Christie, W., editor, *Current Progress in Historical Linguistics*, pages 96–105. North Holland, Amsterdam.

House, A. S. (1961). On vowel duration in English. *The Journal of the Acoustical Society of America*, 33(9):1174–1178.

House, A. S. and Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, 25(1):105–113.

Hyman, L. (2019). *A theory of phonological weight*. De Gruyter Mouton.

Iverson, G. K. and Salmons, J. C. (1995). Aspiration and laryngeal representation in Germanic. *Phonology*, 12(3):369–396.

Jacewicz, E., Fox, R. A., and Lyle, S. (2009). Variation in stop consonant voicing in two regional varieties of american english. *Journal of the International Phonetic Association*, 39(3):313.

Javkin, H. (1977). *Phonetic universals and phonological change*. PhD thesis, U.C. Berkeley.

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

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

Jurafsky, D., Bell, A., Fosler-Lussier, E., Girand, C., and Raymond, W. (1998). Reduction of English function words in switchboard. In *Fifth International Conference on Spoken Language Processing*.

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

Katsika, A. (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*. Routledge, London.

Keating, P. A. (1979). *A phonetic study of a voicing contrast in Polish*. PhD thesis, Brown University.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, 60(2):286–319.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. In Fromkin, V. A., editor, *Phonetic Linguistics: Essays in honor of Peter Ladefoged*, pages 115–132. Academic Press, Orlando.

Kessinger, R. H. and Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, 25(2):143–168.

Kim, C.-W. (1970). A theory of aspiration. *Phonetica*, 21(2):107–116.

Kim, H. and Cole, J. (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, S. and Cho, T. (2012). Prosodic strengthening in the articulation of English /æ/. *Studies in Phonetics, Phonology and Morphology*, 18(2):321–337.

Kingston, J. and Diehl, R. L. (1994). Phonetic knowledge. *Language*, 70(3):419–454.

Klatt, D. H. (1973). Interaction between two factors that influence vowel duration. *The Journal of the Acoustical Society of America*, 54(4):1102–1104.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59(5):1208–1221.

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

Ko, E.-S. (2018). Asymmetric effects of speaking rate on the vowel/consonant ratio conditioned by coda voicing in english. *Phonetics and Speech Sciences*, 10(2):45–50.

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

Kohler, K. J. (1984). Phonetic explanation in phonology: the feature fortis/lenis. *Phonetica*, 41(3):150–174.

Kozhevnikov, V. A. and Chistovich, L. A. (1965). *Speech: Articulation and perception*. Nauka.

Krause, S. E. (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(4):990–995.

Kristoffersen, G. (2000). *The phonology of Norwegian*. Oxford University Press on Demand, Oxford.

Krivokapić, J. ((accepted)). *Prosodic Theory and Practice*, chapter Prosody in Articulatory Phonology. MIT Press. In press Cambridge, MA.

Kulikov, V. (2012). *Voicing and voice assimilation in Russian stops*. PhD thesis, University of Iowa.

Laeufer, C. (1992). Patterns of voicing-conditioned vowel duration in French and English. *Journal of Phonetics*, 20(4):411–440.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *The Journal of the Acoustical Society of America*, 51(6B):2018–2024.

Levy, R. and Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In Scholkopf, B., Platt, J., and Hoffman, T., editors, *Advances in neural information processing systems*, pages 849–856. MIT Press.

Lindblom, B. and Rapp, K. (1971). Reexamining the compensatory adjustment of vowel duration in swedish words. *Stockholm, KTH, Speech Transmission Laboratory Quarterly Progress and Status Report*, 4:19–25.

Lindblom, B. and Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *The Journal of the Acoustical Society of America*, 42(4):830–843.

Lisker, L. (1957a). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, 33(1):42–49.

Lisker, L. (1957b). Minimal cues for separating /w,r,l,y/ in intervocalic position. *Word*, 13(2):256–267.

Lisker, L. (1974). On "explaining" vowel duration variation. *Glossa*, 8(2):233–246.

Lisker, L. (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, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, 29(1):3–11.

Luce, P. A. and Charles-Luce, J. (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(6):1949–1957.

Mack, M. (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *The Journal of the Acoustical Society of America*, 71(1):173–178.

Maddieson, I. (1985). Phonetic cues to syllabification. *Phonetic linguistics: Essays in honor of Peter Ladefoged*, pages 203–221.

Malécot, A. (1968). The force of articulation of American stops and fricatives as a function of position. *Phonetica*, 18(2):95–102.

Massaro, D. W. and Cohen, M. M. (1983). Consonant/vowel ratio: An improbable cue in speech. *Attention, Perception, & Psychophysics*, 33(5):501–505.

Miller, J. L. (1981). chapter Effects of speaking rate on segmental distinctions, pages 39–74. Routledge.

Miller, J. L. and Baer, T. (1983). Some effects of speaking rate on the production of /b/ and /w/. *The Journal of the Acoustical Society of America*, 73(5):1751–1755.

Miller, J. L., Green, K. P., and Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, 43(1-3):106–115.

Miller, J. L., Grosjean, F., and Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*, 41(4):215–225.

Miller, J. L. and Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, 46(6):505–512.

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

Munhall, K., Fowler, C., Hawkins, S., and Saltzman, E. (1992). "compensatory shortening" in monosyllables of spoken english. *Journal of Phonetics*, 20(2):225–239.

Nagao, K. and de Jong, K. J. (2007). Perceptual rate normalization in naturally produced rate-varied speech. *The Journal of the Acoustical Society of America*, 121(5):2882–2898.

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

Nittrouer, S. (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(4):1777–1790.

O'Dell, M. and Nieminen, T. (1999). Coupled oscillator model of speech rhythm. In *Proceedings of the XIVth international congress of phonetic sciences*, volume 2, pages 1075–1078.

Ohala, J. J. (1981). The listener as a source of sound change. In Masek, C. S., Hendrick, R. A., and Miller, M. F., editors, *Parasession on Language and Behavior*, pages 178–203. Chicago Linguistics Society, Chicago.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In MacNeilage, P. F., editor, *The production of speech*, pages 189–216. Springer, New York.

Ohala, J. J. (2011). Accommodation to the aerodynamic voicing constraint and its phonological relevance. In *Proceedings of the 15th International Conference of Phonetic Sciences*, pages 64–67.

O'Kane, D. (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. K. (1973). The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America*, 54(5):1235–1247.

Peterson, G. E. and Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America*, 32(6):693–703.

Pickett, J. M. and Decker, L. R. (1960). Time factors in perception of a double consonant. *Language and Speech*, 3(1):11–17.

Pike, K. L. (1945). *The Intonation of American English*. University of Michigan Academic Press.

Pind, J. (1995). Speaking rate, voice-onset time, and quantity: The search for higher-order invariants for two icelandic speech cues. *Perception & Psychophysics*, 57(3):291–304.

Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, 118(4):2561–2569.

Port, R. F. (1976). *The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words*. PhD thesis.

Port, R. 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, R. F. (1981). Linguistic timing factors in combination. *The Journal of the Acoustical Society of America*, 69(1):262–274.

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

Port, R. F. and Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics*, 32(2):141–152.

Port, R. F., Dalby, J., and O'Dell, M. (1987). Evidence for mora timing in japanese. *The Journal of the Acoustical Society of America*, 81(5):1574–1585.

Priva, U. C. (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, pages 43–48.

Priva, U. C. (2017). Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition*, 160:27–34.

Pycha, A. and Dahan, D. (2016). Differences in coda voicing trigger changes in gestural timing: A test case from the American English diphthong /ai/. *Journal of phonetics*, 56:15–37.

Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *The Journal of the Acoustical Society of America*, 123(2):1104–1113.

Raphael, L. 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(4B):1296–1303.

Raphael, L. J. (1975). The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English. *Journal of Phonetics*, 3(1):25–33.

Raphael, L. J. (1981). Durations and contexts as cues to word-final cognate opposition in English. *Phonetica*, 38(1-3):126–147.

Raphael, L. J., Dorman, M. F., Freeman, F., and Tobin, C. (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(3):389–400.

Repp, B. H. and Williams, D. R. (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(2):445–457.

Revoile, S., Pickett, J., Holden, L. D., and Talkin, D. (1982). Acoustic cues to final stop voicing for impaired- and normal- hearing listeners. *The Journal of the Acoustical Society of America*, 72(4):1145–1154.

Saltzman, E., Nam, H., Krivokapic, J., and Goldstein, L. (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*, pages 175–184.

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

Schwartz, G. (2010). Phonology in the speech signal-Unifying cue and prosodic licensing. *Poznań Studies in Contemporary Linguistics*, 46(4):499–518.

Selkirk, E. (1982). *The structure of phonological representations Part 2*, chapter The syllable, pages 337–383. Dordrecht: Foris.

Sharf, D. J. (1962). Duration of post-stress intervocalic stops and preceding vowels. *Language and speech*, 5(1):26–30.

Sharf, D. J. (1964). Vowel duration in whispered and in normal speech. *Language and Speech*, 7(2):89–97.

Smith, B. L. (2002). Effects of speaking rate on temporal patterns of English. *Phonetica*, 59(4):232–244.

Solé, M.-J. (2007). *Experimental Approaches to Phonology*, chapter Controlled and mechanical properties in speech, pages 302–321. Oxford University Press.

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

Stevens, K. N. and House, A. S. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research*, 6(2):111–128.

Summerfield, A. Q. and Haggard, M. P. (1972). Articulatory rate versus acoustical invariants in speech perception. *The Journal of the Acoustical Society of America*, 52(1A):113–113.

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

Sweet, H. (1880). *A Handbook of Phonetics*. Clarendon Press Series. MacMillan and Co., London.

Tanner, J., Sonderegger, M., Stuart-Smith, J., and Consortium, S. D. (2019). Vowel duration and the voicing effect across English dialects. *Toronto Working Papers in Linguistics*, 41(1).

Toscano, J. C. and McMurray, B. (2015). The time-course of speaking rate compensation: Effects of sentential rate and vowel length on voicing judgments. *Language, Cognition and Neuroscience*, 30(5):529–543.

Treiman, R., Straub, K., and Laver, P. (1994). Syllabification of bisyllabic nonwords: Evidence from short-term memory errors. *Language and Speech*, 37(1):45–59.

Turk, A. E. and Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, 35(4):445–472.

Umeda, N. (1975). Vowel duration in American English. *The Journal of the Acoustical Society of America*, 58(2):434–445.

Van Heuven, W. J., Mandera, P., Keuleers, E., and Brysbaert, M. (2014). Subtlex-uk: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, 67(6):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(3):847–863.

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

Verbrugge, R. R. and Isenberg, D. (1978). Syllable timing and vowel perception. *The Journal of the Acoustical Society of America*, 63(S1):S4–S4.

Viswanathan, N., Olmstead, A. J., and Aivar, M. P. (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(2):436–452.

Volaitis, L. E. and Miller, J. L. (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, T. and Parker, F. (1981). Vowel length and voicing in a following consonant. *Journal of Phonetics*, 9(3):305–308.

Wardrip-Fruin, C. (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(1):187–195.

Weismer, G. (1979). Sensitivity of voice-onset time (vot) measures to certain segmental features in speech production. *Journal of Phonetics*, 7(2):197–204.

Wells, J. C. (1982). *Accents of English*, volume 1. Cambridge University Press, Cambridge.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91(3):1707–1717.

Table 1

Vowel Duration Model for Full Stops

	Estimate	Std. Error	estimated df	t-value	p-value
Intercept	-2.25E-02	4.06E-02	1.84E+02	-0.554	0.580
speaking rate	1.10E+00	2.29E-02	1.30E+02	48.292	<.001
phonetic voicing	-1.58E-02	6.96E-03	1.17E+04	-2.267	0.023
word frequency	-6.32E-02	2.17E-02	2.47E+02	-2.916	<.005
Vowel Class	-1.81E-01	2.04E-02	2.47E+02	-8.861	<.001
phonemic voicing	-1.09E-01	3.50E-02	1.95E+02	-3.122	<.005
Consonant Duration	-2.91E-01	9.41E-03	1.18E+04	-30.946	<.001
Phrasal Position	9.06E-02	3.86E-03	1.17E+04	23.434	<.001
speaking rate :Vowel Class	9.18E-02	2.00E-02	8.87E+01	4.581	<.001
speaking rate :phonemic voicing	1.78E-01	2.26E-02	1.28E+02	7.851	<.001
speaking rate :Consonant Duration	6.48E-02	1.21E-02	9.44E+03	5.35	<.001
word frequency:phonemic voicing	-4.53E-02	2.01E-02	2.40E+02	-2.251	0.025
phonemic voicing:Consonant Duration	-6.57E-02	9.06E-03	1.15E+04	-7.248	<.001

Table 2

Vowel Duration Model for Full Stops in tokens of longest absolute duration

	Estimate	Std. Error	estimated df	t-value	p-value
(Intercept)	0.66267	0.04125	105.06663	16.066	<.001
Vowel Class	-0.0678	0.03412	121.06521	-1.987	0.049
phonemic voicing	0.09313	0.03376	109.00637	2.758	<.01
Consonant Duration	0.18751	0.02447	1023.41315	7.664	<.001

Table 3

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)	357.4	26.37	29.66	13.55	3.06e-14
rate.L	334.4	16.83	838.8	19.88	< 2e-16
rate.Q	51.25	16.71	837.4	3.066	0.002
Consonant Duration	-0.219	0.067	845.0	-3.289	0.001
rate.L:C Duration	-0.734	0.128	838.1	-5.716	1.51e-08

Table 4

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

	Estimate	Std. Error	estimated df	t-value	p-value
(Intercept)	334.4	29.55	23.90	11.32	4.38e-11
rate.L	217.9	11.22	940.6	19.42	< 2e-16
Voicing	54.25	28.37	6.23	1.913	0.103
Consonant Duration	-0.313	0.065	849.1	-4.926	1.01e-06
rate.L:Voicing	36.65	13.48	836.78	2.719	0.007

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)	157.2	8.154	7.268	19.27	1.66e-07
rate.L	87.50	6.791	829.4	12.88	< 2e-16
voicing	-49.54	9.865	4.024	-5.022	0.007
manner	37.80	9.843	3.988	3.332	0.029
rate.L:voicing	-41.84	9.781	829.9	-4.278	2.11e-05
rate.L:manner	22.34	9.657	829.3	2.313	0.021
rate.Q:manner	-21.11	9.651	829.0	-2.187	0.029

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)	-38.55	10.62	7.726	-3.629	0.007
C duration difference	-0.230	0.096	387.2	-2.401	0.017
manner	-38.10	13.35	5.338	-2.855	0.033
C diff:manner:rate.L	-0.585	0.296	381.6	-1.975	0.049

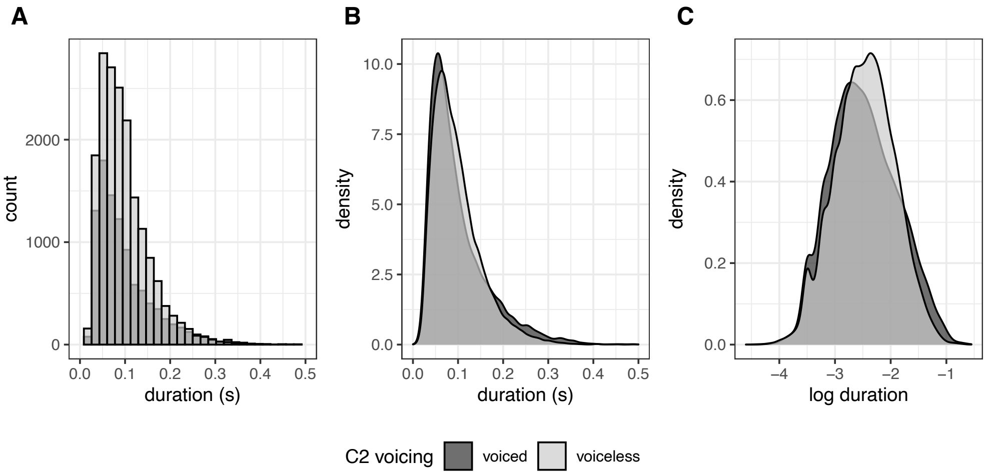


Figure 1

All CVC vowel durations

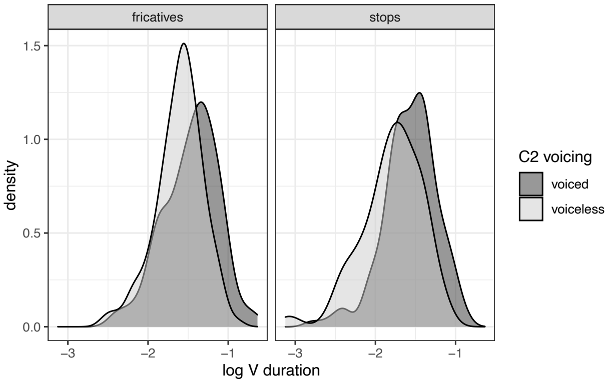


Figure 2

Vowel durations for the longest tokens: content words below the median speaking rate, below the median word frequency, only inherently Long vowels, occurring phrase-finally (878 tokens)

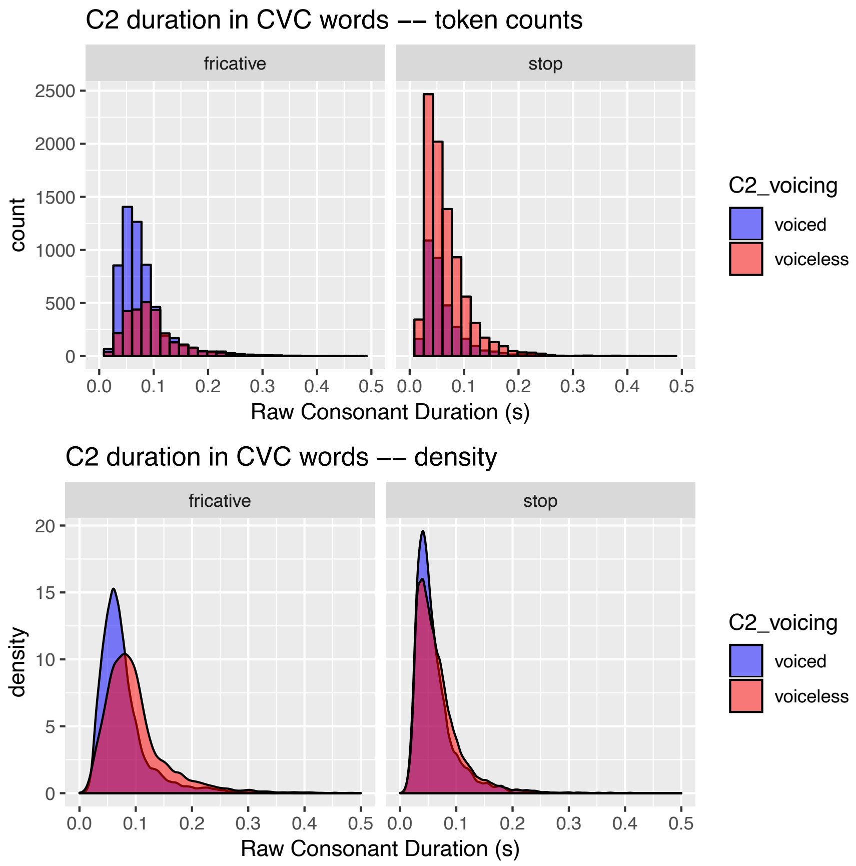


Figure 3

Obstruent durations by fricative and stop (flaps excluded).

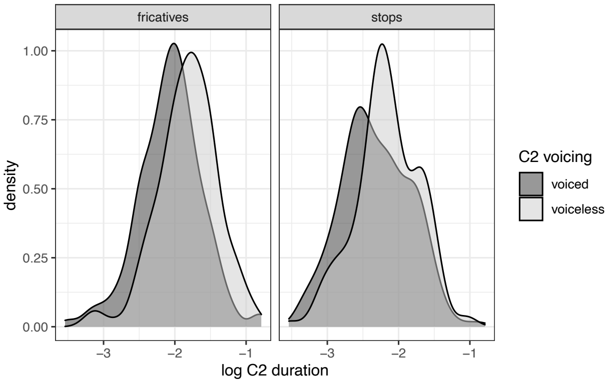


Figure 4

Consonant duration for the subset of longest tokens: content words below the median speaking rate, below the median word frequency, containing the Long vowel class, occurring pre-pausally

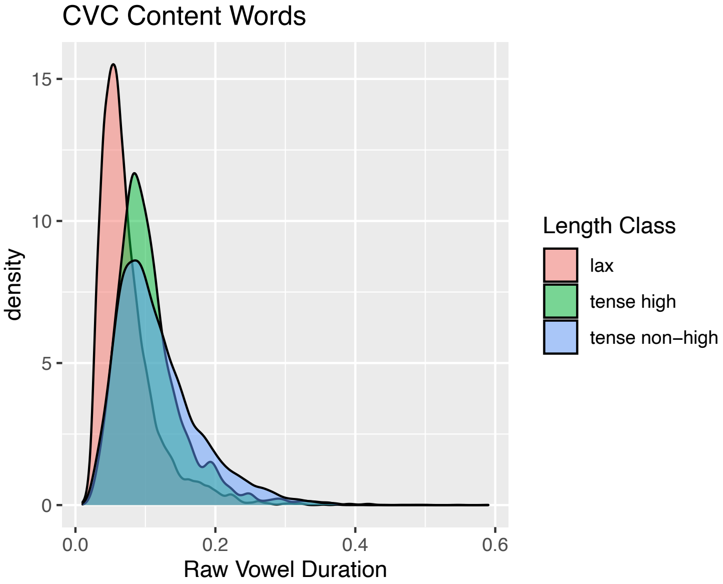


Figure 5

Density distributions of vowel duration for CVC content words in the Buckeye Corpus. Vowels are divided into 3 groups based on predicted duration: lax (ɛ,ə,ʊ,ɪ), tense high (i,e,u), and tense non-high (ɑ,o,æ,ɔ,ɑɪ,ɑʊ)

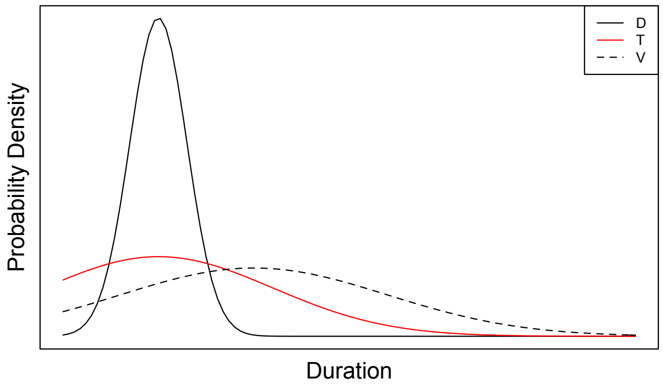


Figure 6

Probability densities for: voiced obstruent (red); voiceless obstruent (blue); vowel (black dashed)

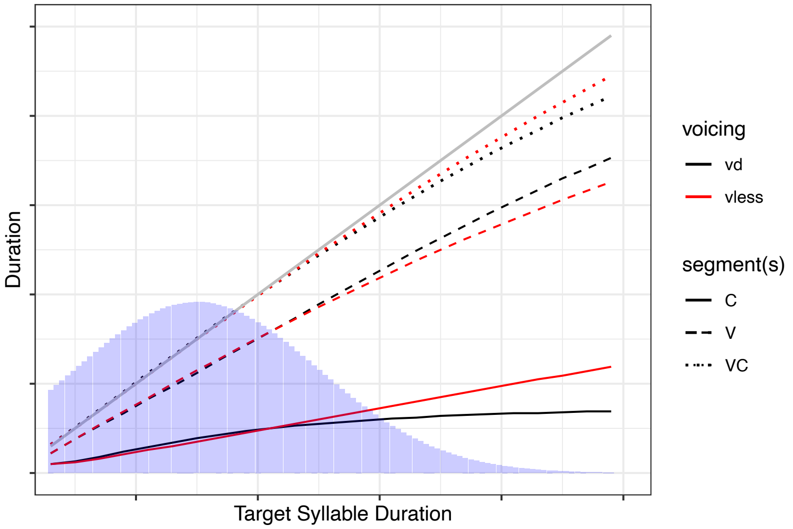
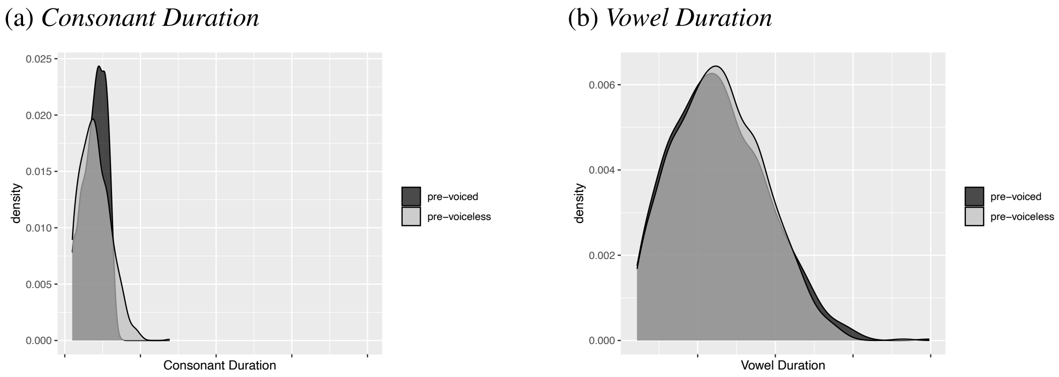


Figure 7

Competing Constraints Model of segment duration as a function of target syllable duration and segment elasticity. Actual and target syllable duration are equal along the solid gray line. Vertical blue lines are estimates of the distribution of durations/rates in the Buckeye Corpus (used for the corpus simulation).


(a) *Consonant Duration*

(b) *Vowel Duration*

Figure 8

Corpus Simulation. Target syllable duration randomly sampled from a Normal distribution shown in Fig. 7.

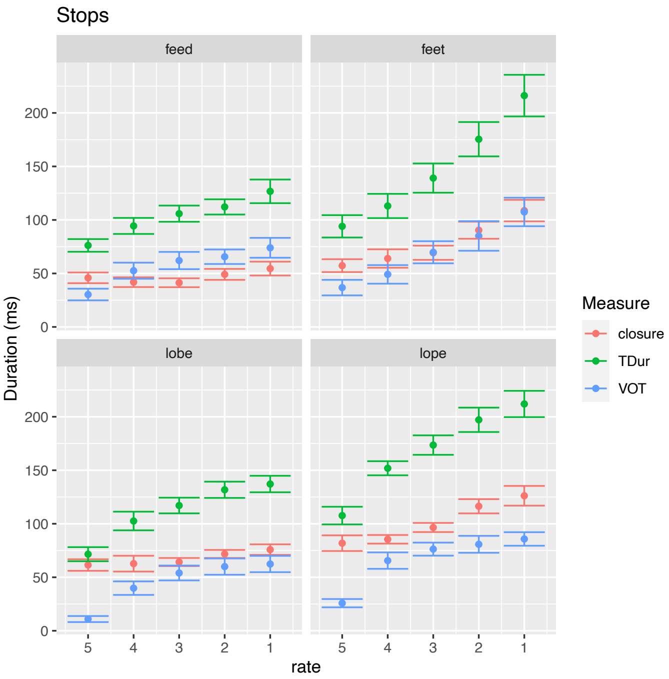


Figure 9

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

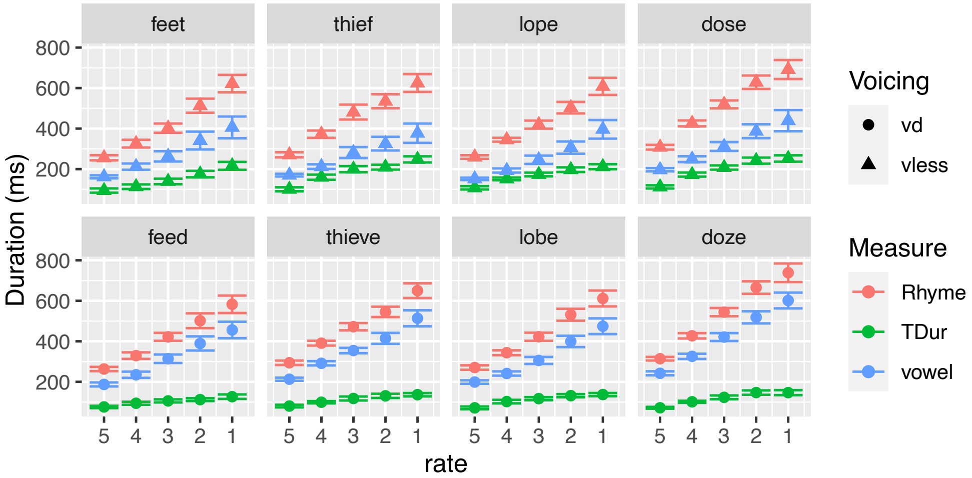


Figure 10

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

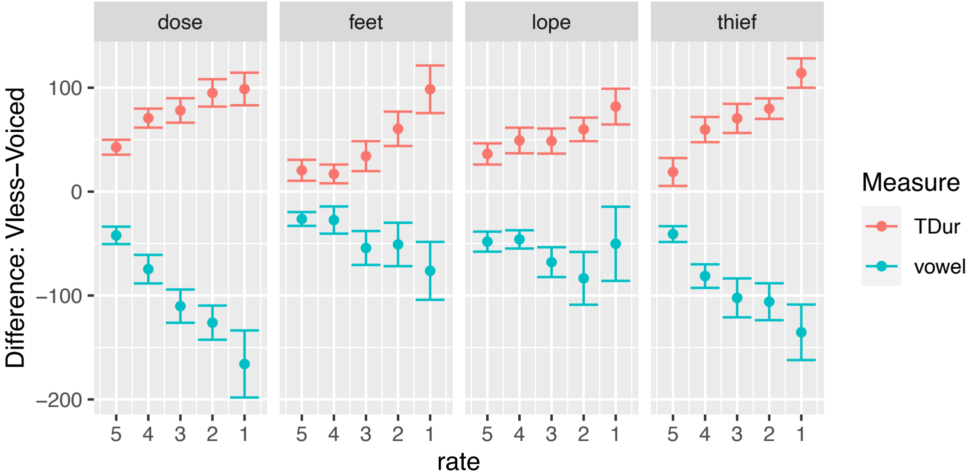
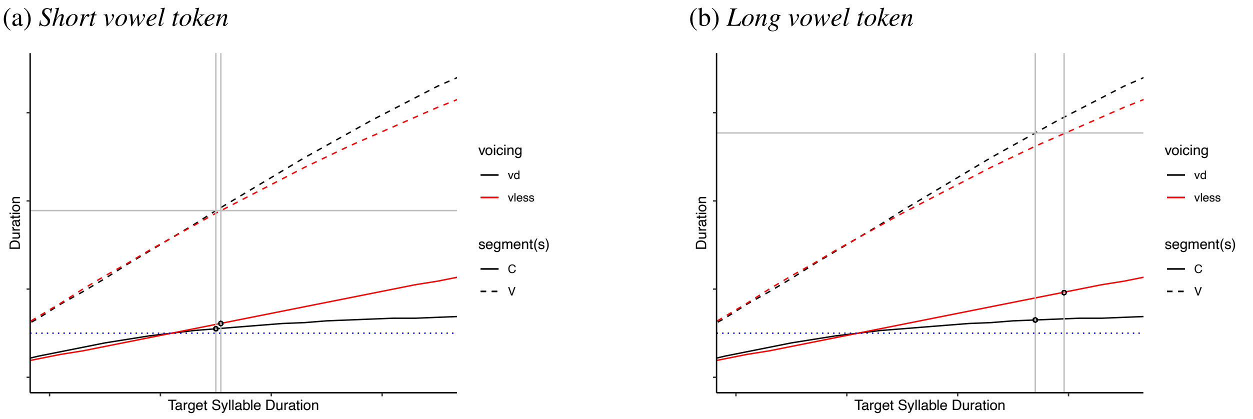


Figure 11

Duration differences between consonants and preceding vowels as a function of repetition rate. Means and standard error bars. change colors: tdur green; vowel blue


(a) *Short vowel token*

(b) *Long vowel token*

Figure 12

Compensation Simulation: Observed vowel duration is marked by the gray horizontal line in both figures. Gray lines intersect expected target syllable duration (speaking rate), and expected stop duration. Left line: voiced stop coda; Right line: voiceless stop coda. The dotted blue line indicates the actual duration of the following stop.

Appendix A

Word Lists

7.0.1 CV Stop

Voiced (4106 tokens; 96 unique words): bad, bag, bed, big, bob, cab, cad, cod, code, could, cub, dab, dad, dead, did, died, dig, dodd, dog, dude, fed, feed, fog, food, gig, god, good, guide, had, he'd, head, hid, hide, hood, how'd, hub, hug, hyde, jed, jedd, job, kid, knob, lab, lag, laid, lead, league, led, leg, lid, lied, life, load, loud, mad, made, maid, med, meg, mid, mud, need, paid, pig, read, red, rhode, rid, ride, road, rob, rode, rub, sad, said, she'd, shed, should, showed, side, sued, tag, ted, they'd, tied, todd, tub, tube, wade, we'd, web, weed, wide, would, you'd

Voiceless (14796 tokens; 183 unique words): back, bat, beat, beep, bet, bike, bit, bite, boat, book, bought, buck, but, butt, cake, cap, cape, cat, caught, chalk, cheap, check, chick, chip, coke, cook, cop, cope, cup, cut, date, debt, deck, deep, dip, dot, doubt, duck, duke, fake, fat, feet, fight, fit, folk, foot, fought, fuck, gap, gate, get, got, gut, hat, hate, heat, heck, height, hick, hip, hit, hook, hop, hope, hot, hype, jack, jeep, jet, jock, joke, kat, keep, kick, knit, know, lack, lake, lap, late, let, light, like, lock, look, lot, luck, luke, mac, make, map, mat, meet, met, might, mike, mock, nap, neat, neck, net, night, nope, nose, not, note, nut, pack, pat, peek, pet, pete, pick, pipe, poke, pop, pope, pot, psych, puck, put, rat, rate, rec, right, rock, rope, route, sake, sat, seat, set, shake, shape, sheet, ship, shit, shock, shoot, shop, shot, shut, sick, sight, sit, site, soap, soup, suit, take, talk, tap, tape, taught, tech, that, thick, this, thought, tight, tip, took, top, type, vet, vote, wait, wake, week, weight, wet, whack, what, whip, white, wick, woke, wreck, wright, write, wrote, yet, zip

7.0.2 CV Fricative

Voiced (5666 tokens; 77 unique words). b's, boys, c's, cahs, cause, cave, cheese, choose, chose, cows, d's, days, dies, does, dos, faze, five, gave, gays, give, goes, guys, has, have, hayes, haze, he's, his, how's, jazz, joe's, keys, knees, knows, laws, leave, live, lose, love, move, news, noise, p's, pays, phase, raise, rave, rise, rose, save, says, seas, sees, shave, she's, shoes, shows, size, so's, t's, taj, these, they've, those, ties, toes, toys, twos, use, was, wave, ways, we've, who’s, whose, wise, you’ve

Voiceless (2813 tokens; 80 unique words). base, bash, bass, bath, beef, biff, boss, both, bus, bush, calf, case, cash, chess, chief, choice, cuff, cuss, dose, face, faith, fish, fuss, gas, geese, goose, gosh, guess, half, hash, house, joyce, juice, kiss, knife, las, laugh, lease, less, loose, los, mass, math, mess, mice, miss, moss, mouth, nice, niche, niece, pace, path, peace, piece, piss, push, race, rash, reese, rice, rough, rush, safe, south, teeth, this, thought, tiff, tooth, toss, tough, vice, voice, wash, wife, wish, with, yes, youth

Appendix B

Campbell's Elasticity Model

In Campbell (1992) expansion and compression are relative to mean duration, while standard deviation is used as proxy for elasticity. The model is based on the hypothesis that a single expansion coefficient (ε_k) can be applied to all segments (S_i) within a given syllable (σ_k), expanding each by the same number of standard deviations (κ_i) from its own mean. See Equation (B1).

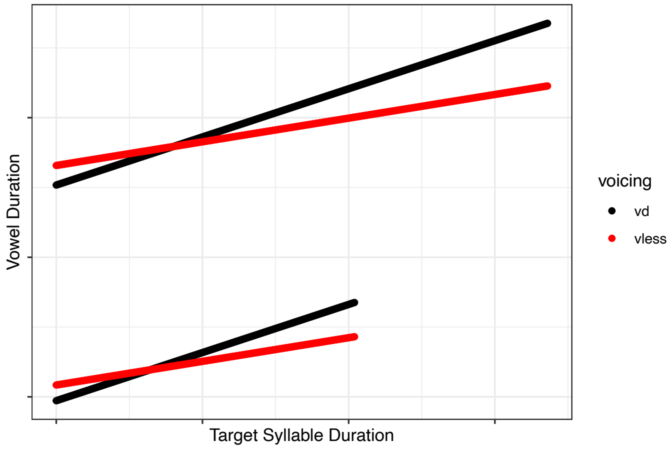
$$S_i = \bar{S}_i + \kappa_i \varepsilon_k \tag{B1}$$

The appropriate expansion coefficient is found by taking the difference between the baseline syllable duration ($\overline{\sigma}$) and the target syllable duration (σ_T), divided by the sum of the elasticities of all segments within the syllable. See Equation (B2).

$$\varepsilon_k(\sigma_T) = \frac{\sigma_T - \bar{\sigma}_k}{\sum\limits_{i \in k} \kappa_i} \tag{B2}$$

For a voiced/voiceless minimal VC syllable pair (VD, VT), with the same target duration, $\varepsilon_{vd} = \frac{\sigma_T-(\bar{V}+\bar{D})}{\kappa_V+\kappa_D}$, and $\varepsilon_{vless} = \frac{\sigma_T-(\bar{V}+\bar{T})}{\kappa_V+\kappa_T}$. Therefore, $\varepsilon_{vless} = \frac{\varepsilon_{vd}(\kappa_V+\kappa_D)+(\bar{D}-\bar{T})}{\kappa_V+\kappa_T}$. Because the elasticity of the voiced obstruent is less than the elasticity of the voiceless obstruent, $\frac{\kappa_v+\kappa_D}{\kappa_v+\kappa_T}$ is less than 1. Therefore, as target syllable duration increases, ε_{vless} increases more slowly than ε_{vd}. However, when target syllable duration is short enough to lead to compression (negative values for ε_{vd}), the opposite relation holds. A higher elasticity means a segment can both lengthen more, and compress more. The difference between the more elastic and the less elastic segment will continue to increase in both directions away from the mean.

The voicing effect, $V_{vd} - V_{vless}$, is given by $\kappa_V \varepsilon_{vd} - \kappa_V \varepsilon_{vless}$, which can be rewritten as $\varepsilon_{vd}\kappa_V(1 - \frac{\kappa_V+\kappa_D}{\kappa_V+\kappa_T}) + \frac{\kappa_V(\bar{D}-\bar{T})}{\kappa_V+\kappa_T}$. ε_{vd}, in turn is a linear function of σ_T. The magnitude of the voicing effect, for any given target syllable duration, can be determined via: $\frac{\sigma_T-(\bar{V}+\bar{D})}{\kappa_V+\kappa_D}\kappa_V(1 - \frac{\kappa_V+\kappa_D}{\kappa_V+\kappa_T}) + \frac{\kappa_V(\bar{D}-\bar{T})}{\kappa_V+\kappa_T}$. As long as ε_{vd} is positive, the difference in duration between pre-voiced and pre-voiceless vowels will increase linearly with target syllable duration, at a rate given by $\kappa_V(\frac{1}{\kappa_v+\kappa_D} - \frac{1}{\kappa_v+\kappa_T})$. When ε_{vd} is negative, the voicing effect reverses at the same rate. A somewhat simplified version of Campbell's model,[34] using only the above equations, was used to generate Fig. B1. The upper and lower sets of lines correspond to inherently longer and shorter vowels, respectively.



[H]

Figure B1
Campbell's Simplified Model: VC syllables.

For a VC syllable, $\sigma = V + C$. If the change in duration for a given segment, S, is denoted by ΔS, then $\sigma_T = \bar{V} + \Delta V + \bar{C} + \Delta C$. For a given target duration (larger than the mean), a larger expansion coefficient is required for the voiced syllable, which has the effect of lengthening the pre-voiced vowel more than the pre-voiceless, and is the source of the voicing effect. More specifically, the vowel in the voiced syllable must be lengthened by precisely the amount necessary to compensate both for the discrepancy between the expansion of the voiceless and voiced obstruent, and for the difference between their mean durations:
$\Delta V_{vd} - \Delta V_{vless} = (\bar{T} - \bar{D}) + \Delta T - \Delta D$.

[34] Target syllable duration was treated as a random variable, ranging over multiples of the baseline syllable duration, rather than being fit to the phonological properties of the syllable.

The difference between inherently long and inherently short vowels is modeled at the syllable level by assigning different mean durations to the two kinds of syllables. The result is that "short" vowels undergo less lengthening, on average, than "long" vowels.[35] This also means that a difference in the magnitude of the voicing effect for shorter versus longer vowels appears only in the aggregate data. Pair-wise comparisons (at the same target duration) between long and short vowel syllables will show no difference in the magnitude of the voicing effect. Note that all results rely on assigning a smaller mean and variance to the voiced obstruent than to the voiceless, even though differences between the two are small to non-existent in the Buckeye Corpus.

[35] It is worth noting that the difference in vowel type cannot be captured at the level of the segment in Campbell's model. The segment-level modeling implicitly requires the ability to lengthen to any degree. A larger expansion coefficient is simply applied to less elastic segments in order to achieve the same length. Thus, not only would short vowels get as long as long vowels, a larger voicing effect would occur in short versus long-vowel syllables because segments would be subjected to a larger ε on average, the opposite of what is observed.

Appendix C

Competing Constraints Model

Constraints in this model are realized as Normally distributed probability densities. The 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 determined by the elasticity of the segment. 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.

The full set of constraints for the competing constraints model is given in (C1), 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. The actual values for the variances, however, were chosen to produce differences large enough to exhibit the desired behavior. Because observed duration variance is not necessarily the same as elasticity, and the variance of the individual probability distributions represent elasticity, and not duration variance, this should not be problematic.

(C1) $P(\frac{C}{V}) \sim \mathscr{N}(\mu = .3, \sigma = .1)$

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

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

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

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

$\frac{V}{\sigma}$: 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), is given as

$$p(y,z,\sigma_T) = p(\frac{C}{V} = \frac{y}{z}) \cdot p(C = y) \cdot p(V = z) \cdot p(\frac{\sigma_T - (y+z)}{\sigma_T}) \tag{C2}$$

To reduce run time, the constraint $(\frac{V}{\sigma})$ is implemented by simply restricting the search space.[36]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 100. 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 (durations are not allowed to fall below 20 ms.). The same 1000 point sample of targets was used for both the voiced and voiceless distributions. The results for VC syllables are given in Section 3.3. See Figures 7 and 8.

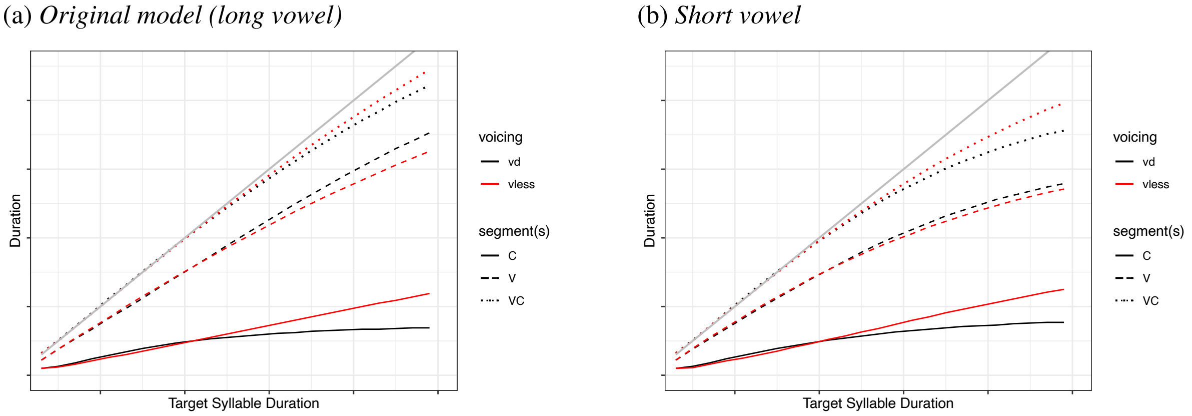
Short Vowels. We assume that target syllable duration is determined by a combination of speaking rate and other prosodic factors, such as phrase-final lengthening (see, e.g., Byrd and Saltzman 2003). In Campbell (1992), shorter vowels were essentially given a smaller range of possible target durations relative to longer vowels. This approach is not without justification, given that the nucleus type is often treated as equivalent to the syllable type, and different syllable types may have their own associated duration ranges. A similar approach could be implemented in the competing constraints model. However, the architecture of our model offers an alternative way to differentiate long and 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 C1 shows the result of reducing the variance of the vowel probability distribution

[36] 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.

($\sigma = 40$: 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. 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, greater target undershoot results in a higher joint probability than lengthening either consonat additionally. 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. 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 C1
Competing Constraints Model

CV syllables. Given that differences in duration between voiced and voiceless obstruents in initial position have been shown to have some effect on the duration of following, tautosyllabic vowels, it should be possible to develop a broadly similar competing constraints model that accounts for these differences. 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 4.2. 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 also lead to greater masking than inherently shorter consonants; an acoustically shorter vowel following a voiceless obstruent than a voiced one, is predicted in the range where voiceless obstruents are loner 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. C2 will cover both the VC and CV cases. The rhyme in the VC case is calculated by adding the durations of the consonant and vowel (assuming no overlap).[37]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[38]. 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.

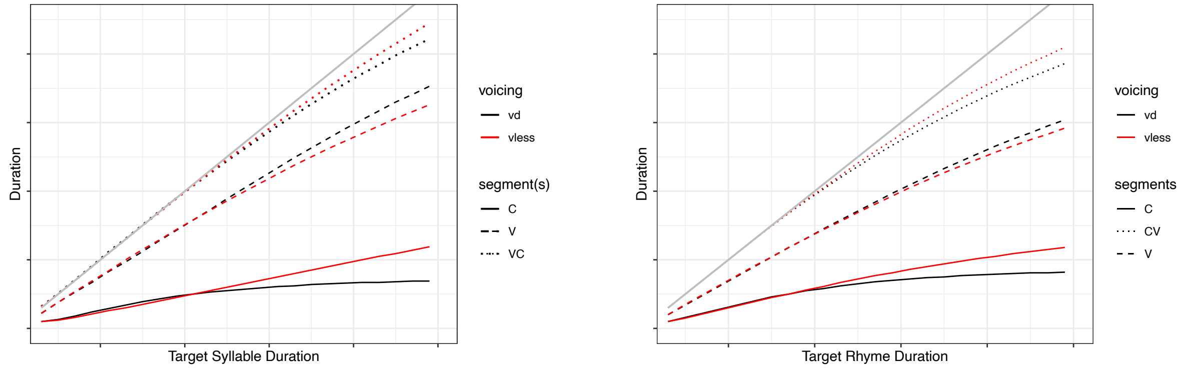
These changes alter the model behavior in the desired ways. See Figure C2. The VC

[37] 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.

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

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 acheive 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.

[H]


(a) *Coda model (original)*

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

Figure C2
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. C2). 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 summarized in Section 5.2.