Influence of coda stop features on perceived vowel duration

Chelsea Sanker[a]*

*Corresponding author: chelsea_sanker@brown.edu, Tel: +1 401 863 5190
[a] Brown University, Providence, 02912, Rhode Island, USA

Abstract
Four experiments tested what cues contribute to English speakers' perception of vowel duration. Listeners categorized the duration of vowels as 'long' or 'short' for stimuli produced with voiced, voiceless, breathy voiced, or voiceless aspirated stop codas. Listeners demonstrated a strong ability to perceive vowel duration, though perception was continuous rather than categorical. There were several interacting factors influencing perceived vowel duration, based on expectations set by the presence of particular codas and also acoustic effects of the coda on the vowel. When the coda was removed, vowels that had been produced before voiced codas were perceived as longer than vowels produced before voiceless codas, though they exhibited the opposite effect when codas were present. Vowels were also perceived as longer when produced before breathy voiced stops, regardless of whether or not the stop was present. The steeper f0 falls associated with voiced codas within these stimuli likely contributed to the longer perceived duration of vowels from this environment; manipulating f0 contours eliminated effects of the original coda on perceived vowel duration. The effects of the production environment on perceived vowel duration suggest a possible perceptual pathway for the voicing effect on vowel duration.

Keywords
Voicing effect; vowel duration; speech perception

Influence of coda stop features on perceived vowel duration

Chelsea Sanker

1.0 Introduction

Vowels are longer before voiced consonants than voiceless consonants in many languages. Explanations have been proposed based on articulation (e.g. Chen, 1970; Halle & Stevens, 1967) and perception (e.g. Kluender, Diehl, & Wright, 1988; Javkin, 1976), but the underlying cause remains in question. The effects of aspiration and breathiness on vowel duration have been studied less, but may help inform the effects of voicing on vowel duration. Both perception and production may play a role, as any perceptually-driven effect is only possible if there are differences that exist based on production.

The experiments presented here probe the possibility of a perceptual source of voicing effects on vowel duration, using English speakers' responses to a listening task in which they categorized a continuum of vowel durations as 'long' or 'short'. The experiments examine what factors other than duration itself influence listeners' perception of vowel duration. The results suggest that coda voicing influences vowel characteristics in ways that contribute to the perception of duration; vowels sound longer when they were produced before voiced stops than when they were produced before voiceless stops. However, this effect interacts with effects of phonological knowledge; listeners partially compensate for the duration expected in the phonological environment, decreasing their number of 'long' responses for vowels which are presented with a following voiced coda.

1.1 Duration and voicing in production

Many studies demonstrate that vowels are longer before voiced consonants than before voiceless consonants, in a range of languages. The effect is often studied in English (e.g. Peterson & Lehiste, 1960; Chen, 1970), but also significant in many other languages, such as German (Fourakis & Iverson, 1984), Dutch (Warner et al., 2004), Marathi, Bengali (Maddieson, 1977), French, Korean (Chen, 1970), and Japanese (Port, Al-Ani, & Maeda, 1980).

There are relatively fewer studies on how the effect of voicing interacts with aspiration and breathiness of coda stops. Within Hindi, Durvasula & Luo (2014) and Maddieson & Gandour (1976) found longer vowels before voiceless aspirated stops and breathy voiced stops than their unaspirated or plain counterparts, though Ohala & Ohala (1992) found no consistent difference. Work on other languages with breathy voiced stops tends to find longer vowels before these stops than before plain voiced stops, though there is some variation by language (Maddieson, 1977).

Languages vary in the degree of the voicing effect, which indicates that even if it is driven by mechanical effects of anatomy or perceptual biases, there is a role of language-specific learning in the implementation of it. Some languages exhibit a much smaller difference than others in vowel duration before voiced and voiceless consonants, though some of the differences in effect size are likely due to variations in experimental design and not just differences by language (Laeufer, 1992); e.g. Keating (1979) found no significant voicing-conditioned duration differences in Polish or Czech, while others have found small but significant differences in Polish (Nowak, 2006: 157) and Czech (Machač & Skarnitzl, 2007) and similarly Mitleb (1984) found no significant effect of coda voicing on vowel duration within Levantine Arabic, while Hussein (1994) found a small but significant effect.

Differences in phonologization of the voicing effect are suggested by how it behaves in contexts of neutralization. In some languages, the duration differences are preserved even in environments where the voicing contrast is eliminated, as with German word-final devoicing (Fourakis & Iverson, 1984), and when the laryngeal contrast is not realized with voicing, as in whispered English (Sharf, 1964). On the other hand, the duration difference is lost in voicing neutralization contexts in other languages, e.g. with final devoicing in Dutch (Warner et al., 2004).

Differences in phonologization are also suggested by language-specific patterns of how the voicing effect interacts with other influences on vowel duration. For example, in Arabic, duration differences due to phonological vowel length are proportionally expanded in stressed syllables, but the absolute differences in duration based on coda voicing are equivalent between stressed and unstressed syllables (de Jong & Zawaydeh, 2002) and between phonologically long and short vowels (Port, Al-Ani, & Maeda, 1980). In contrast, the proportional size of the voicing effect in English is similar in stressed syllables (de Jong, 2004), under phrase-final lengthening (Cooper & Danly, 1981), and across vowels of different inherent duration (Peterson & Lehiste, 1960).

Voicing is associated with a range of characteristics beyond VOT, as discussed in the recent overview by Cho, Whalen, and Docherty (2019), and summarized for coda voicing by Lisker (1986). Some studies have found higher f0 before voiceless consonants than before voiced consonants (e.g. Kohler, 1982), though other studies have not found this difference (e.g. Mohr, 1971; Gruenenfelder & Pisoni, 1980); breathy voicing also decreases f0 (Hombert, Ohala, & Ewan, 1979). The formant transition from the vowel to the consonant closure is shorter before voiceless codas (Summers, 1987; de Jong, 1991). F1 is also higher before voiceless consonants than before voiced consonants, at least for low monophthongs (Summers, 1987), though voicing has the opposite effect within diphthongs (Moreton, 2004). Pycha & Dahan (2016) propose a gestural coordination account of voicing-conditioned duration effects and formant effects; while not central to their account, they observe that coda voicing, phrasal position, and speech rate all have similar effects both on overall vowel duration and on peripherality of the vowel formants. These results might suggest that the F1 effects of voicing are at least partially driven by duration differences rather than being a direct effect of voicing; consistent with this possibility, Sanker (2018) found that Hindi, which has a voicing effect, also exhibits lower F1 in monophthongs (/a/, /i/, /u/) before voiced consonants, while Telugu, which lacks a robust voicing effect, does not exhibit an effect of voicing on F1. Spectral tilt (H1-H2) has also sometimes been observed to be higher in vowels next to voiceless stops than next to voiced stops (Kong et al., 2012; Coleman, 2003; Al-Tamimi & Khattab, 2018); glottalization, with lower spectral tilt, is associated with some voiceless stops, particularly in codas (Chong & Garellek, 2018; Seyfarth & Garellek, 2018; Penney et al., 2018). These acoustic differences likely reflect articulatory differences, but they might also contribute to perception of vowel duration.

1.2 Duration and voicing in perception

Some possible perceptual explanations for the voicing effect have been proposed, though subsequent studies have not provided clear support to establish any of them. Nonetheless, there have been various studies demonstrating that some acoustic characteristics that are influenced by coda voicing can play a role in perceived vowel duration.

Javkin (1976) suggested that vowels are perceived as longer before voiced codas than before voiceless codas because listeners misinterpret the boundary between the vowel and the following consonant when voicing is maintained throughout. Lengthening due to ambiguous boundaries would likely be similarly apparent as an effect of following consonants and preceding consonants; some studies have found lengthening following voiced consonants (Mohr, 1971; Port, Al-Ani, & Maeda, 1980), though others have not (e.g. Peterson & Lehiste, 1960).

Kluender, Diehl, & Wright (1988) propose that vowels are perceived as longer when they are next to shorter consonants, due to having longer relative duration, and that the shorter duration of voiced obstruents thus leads to greater perceived duration of preceding vowels. However, Fowler (1992) tested this explanation, and found that longer stop closures actually increased the perceived duration of preceding vowels, rather than decreasing it.

Paralleling effects of codas on vowels in production, acoustic characteristics of vowels also contribute to listeners' identifications of voicing of consonants. Vowel duration is used as a cue for identifying coda consonants, with longer durations increasing listeners' perception that codas are voiced (Raphael, 1972; Warner et al., 2004). Lower f0 within the vowel and particularly at its end similarly increases listeners' perception that a following stop is voiced (Gruenenfelder & Pisoni, 1980; Kohler, 1985), as does lower F1 (Summers 1988; Benkí 2001). Longer transitions from the vowel to the consonant closure also result in more identifications of coda consonants as voiced (Stevens & Klatt, 1974; Benkí, 2001).

Studies examining possible perceptual sources of the voicing effect tend to focus on how characteristics of the coda influence perception of the vowel. There are fewer studies that test how these acoustic characteristics might influence perceived duration of vowels. In production, lower vowels (i.e. vowels with a higher F1) are longer than higher vowels; in perception, higher vowels are identified as longer, likely as a compensatory effect (Gussenhoven, 2004). However, such studies compare across vowel categories; it is not clear if F1 would have the same effect within categories. Some work in languages with contrastive vowel length have found perceptual effects of formant structure on perceived duration, even within contrasts that are primarily duration-based; more peripheral and diphthongal vowels are more likely to be identified as long (Hadding-Koch & Abrahamson, 1964; Abrahamson & Ren, 1990).

In this paper, I use splicing to separate effects of voicing within the coda from effects of characteristics within the vowel, as driven by the production environment. Study 1 tests how having a voiced or voiceless coda influences perceived duration of a vowel. Studies 2a and 2b test how the original production environment of a vowel influences its perceived duration, reflecting perceptual effects of some of the acoustic differences in vowels produced before voiced and voiceless codas. Studies 2a and 2b also test whether having voicing from just the closure portion of a /d/ spliced onto the end of a vowel increases perceived duration, testing Javkin's (1976) proposal about the indistinctness of segment boundaries.

There is a larger body of work examining the effect of f0 on perceived vowel duration. Higher f0 increases the perceived duration of vowels (Yu, 2010; Gussenhoven & Zhou, 2013) and f0 contours similarly increase perceived vowel duration (e.g. Yu, 2010; Lehiste, 1976), so f0 drops caused by voicing could potentially increase the perceived duration of the preceding vowel. However, the effect of f0 contour varies based on environment, and is clearest in monosyllabic words in isolation (van Dommelen, 1993). In Study 3, I test the effects of f0 contours and how they interact with effects of the coda environment.

How duration perception is tested might influence results. Many studies with English speakers ask them to indicate which of two paired vowels sounded longer (e.g. Lehiste, 1976; Fowler, 1992); such juxtaposition may elicit different processing than decisions in isolation. Javkin's (1976) task matching the duration of a tone to a simultaneous vowel could similarly facilitate comparison rather than categorization. With speakers of languages with vowel length contrasts, duration perception can be indicated using the vowel length categories (e.g. van Dommelen, 1993). English listeners can also categorize vowel duration, though few studies ask them to do so; Yu (2010) had listeners rate duration along a 7-point scale, and Keating (1985) had English listeners use 'long' and short' to categorize Czech vowels. Many of these studies use synthetic stimuli, which may also influence how listeners perceive with the stimuli, based on their limited linguistic naturalness. Binary categorization of vowel duration, paralleling languages with contrastive length, may provide the most linguistically natural way of testing duration perception, and using manipulations of natural speech rather than purely synthetic speech is likely to further increase naturalness.

2.0 Experiment 1

2.1 Recordings

Stimuli were made from recordings of a male Hindi speaker reading VC nonce words, elicited individually in randomized order with PsychoPy (Pierce, 2007) and recorded in a sound attenuated booth with a stand-mounted Blue Yeti microphone in the Audacity software program and digitized at a 44.1 kHz sampling rate with 16-bit quantization. In the items selected for use as stimuli, all stops included a release burst, and all voiced and breathy voiced stops contained voicing throughout the closure. Table 1 presents a summary of the acoustic characteristics of vowels in the recordings which the stimuli were made from.

Table 1 ***Acoustic characteristics of the vowels in the stimuli, by the following consonant's features.***

	f0 mean	f0 Δ (Q1-Q4)	Intensity Δ (Q1-Q4)	F1 (z-scored)	Spectral tilt	HNR	Jitter
Voiced	146.9 Hz	-17.0 Hz	-0.14 dB	-0.29	4.28 dB	6.56	0.011
Voiceless	148.5 Hz	-8.1 Hz	-0.63 dB	0.013	4.69 dB	4.69	0.013
Breathy Voiced	147.6 Hz	-12.4 Hz	-0.01 dB	-0.11	5.12 dB	6.78	0.015
Voiceless Aspirated	149.0 Hz	-14.8 Hz	-0.76 dB	0.080	4.30 dB	4.12	0.016

While these recordings did not include enough tokens for differences to be significant, there were several patterns of differences in vowel qualities based on the following consonant, consistent with previous work on effects of coda consonants on vowels. As measured from the first quarter of the vowel to the last quarter of the vowel, there were larger f0 drops and less of an intensity drop in vowels followed by voiced stops than in vowels followed by voiceless stops. Measured across the whole vowel, there was lower F1 before voiced stops than before voiceless stops and a higher harmonics-to-noise ratio. Jitter was only slightly higher before voiceless stops than before voiced stops. The spectral tilt was greater before breathy voiced stops than in other environments, reflecting breathiness within the vowel. Some of these characteristics might be contributing to perceived duration.

Hindi productions were selected both to allow a comparison of voiced and breathy voiced stops and to provide more control over vowel duration. The comparison of voiced and breathy voiced stops provides additional information on possible acoustic influences on perceived duration; the effects of breathiness may help elucidate effects of voicing, based on exhibiting a different set of characteristics within the coda and within the preceding vowel.

While vowels in Hindi are longer before voiced codas than before voiceless codas, the effect is smaller than it is in English, which makes it possible to select base stimuli in which duration does not differ greatly based on the voicing environment, reducing possible effects of duration differences in the original productions that could carry over into differences in perception. Moreover, Hindi has contrastive vowel length, which is largely realized with duration. To minimize possible differences in the effects of adding glottal cycles and removing glottal cycles, while still starting with naturally produced items, all of the vowels in the nonce words were long. The naturally produced long vowels were long enough that most stimuli could be made by removing glottal cycles; the only items that contained copied glottal cycles were the longest duration step for /a/ before voiceless codas and the two longest duration steps for /i/ and /u/ before voiceless codas. These differences in the manipulations based on vowel quality and coda voicing could potentially cause differences in perception, which should be apparent only at these long durations for which the manipulations differed.

It is worth considering how English speakers approach this categorization task. The perceptual assimilation model (PAM) suggests that success in discriminating between unfamiliar sounds is often the result of mapping the sounds onto contrastive native categories (Best, 1995). Listeners tend to focus on cues that are salient within their native language (Choi, Kim, & Cho, 2016; Chang, 2018), and discrimination of L2 contrasts is predicted by how well the contrasts align with the phonotactics and phonemic inventory of the listener's native language (Davidson, 2011). It is not clear whether the duration contrast that listeners were asked to make would map onto a native English category. It could potentially align with stress, for which duration is a major component, which would likely result in responses influenced by other correlates of English stress, such as f0. It also could potentially align with vowel duration as influenced by coda voicing; this would predict that codas will influence categorizations in the same way as vowel duration, with voiced codas and long duration of the vowel both increasing 'long' responses. Mapping to stress or to coda voicing would predict a largely categorical pattern of responses. However, it is possible that people can perceive duration without mapping it onto a native phonological contrast, as many studies have demonstrated that listeners can accurately evaluate duration on this time scale even for non-linguistic sounds (e.g. Abel, 1972; Rammsayer & Lima, 1991); this lack of mapping would predict a non-categorical pattern of responses.

2.2 Experiment Design

24 native speakers of American English (8 male; mean age 20.2, age range: 18-24) participated in the study. Each participant completed two duration perception experiments; they were randomly assigned to two of the four experiments, and the order of the tasks was randomly determined. None of them had any linguistic training. They heard VC syllables presented over headphones and identified the vowel in each as 'short' or 'long', indicated with the arrow keys on a computer keyboard. The experiment was run in PsychoPy.

Participants were instructed to make decisions about vowel duration and not about vowel quality. The instructions were given verbally: "You are going to hear nonsense words and categorize the vowel in each one as being long or short in duration. This is purely about how long the vowel lasts; a short vowel like [æ̆] or a long vowel like [æː]. This is not a contrastive characteristic of English, so it is up to you to decide what feels like the right divide for you."[1] Then were then asked whether the task was clear, and were invited to ask any questions to clarify; all of them confirmed their understanding. They did not receive any feedback to train them on a particular duration divide.

Stimuli included 8 plosive codas (dental and velar, with each of the four combinations of voice and spread glottis that are present within Hindi stops) and 3 vowel qualities (/a, i, u/). Vowel duration was manipulated to create a 10-step continuum (from ~129 ms to ~252 ms); the difference in duration between each duration step was approximately 13.7 ms.

This produced a total of 240 items, which were heard a single time by each listener. Items were blocked by vowel quality, and randomized within blocks. Across participants, the order of blocks was balanced, and the arrow keys associated with 'long' and 'short' were balanced. A new trial began 200 ms after a response was given.

Statistical results are based on logistic regression using a binomial distribution, calculated in R version 3.4.3 (R Core Team 2017) with the lme4 (version 1.1-7) package (Bates et al., 2015). *p*-values were calculated by the lmerTest (version 2.0-33) package (Kuznetsova et al., 2015).[2] Responses with latencies shorter than 250 ms or longer than 5 s were excluded from analysis (< 1% of the data). Most crucially, this removed the most extreme fast responses that occurred

[1] The IPA notation here indicates that the experimenter provided an example of a phonetically short vowel and a phonetically long vowel; participants had no experience with IPA and instructions were given verbally to ensure that duration could be illustrated for them in this way.
[2] Tables of the response data are included as supplementary data; the audio stimuli are not included, as the speaker did not consent to having recordings of his voice publicly distributed. The R syntax for each of the statistical analyses is given in the Appendix.

before the listener had heard the majority of the vowel. This exclusion also removed trials with the most extreme response times in the long right tail, which likely reflected instances of distraction and responses given beyond when listeners were likely to have retained acoustic details of the stimulus. Because of the skewed distribution of response times, typical outlier exclusions were unsuitable, because they found a large number of long latency outliers, while not consistently excluding responses that were given before the listener heard the vowel.

2.3 Hypotheses

Hypothesis 1: The voicing of the coda might influence the perceived duration of a preceding vowel, based either on English speakers' awareness of a phonologically predictable relationship, or based on acoustic effects that influence perceived duration.

(a) There might be no effect of coda voicing on perceived duration, either indicating that coda voicing does not influence perceived vowel duration, or indicating that there are opposing effects that balance each other out.

(b) Listeners might compensate for the vowel duration that is phonologically expected in the environment of the given coda, and identify vowels as being longer before voiceless codas. They expect longer vowels before voiced codas, so a vowel of a particular duration is shorter relative to expectations before a voiced coda than before a voiceless coda.

(c) Consonant voicing might produce acoustic characteristics in preceding vowels that increase the perceived duration of those vowels, which would result in listeners more frequently identifying vowels as long when they were produced before voiced codas.

Hypothesis 2: Breathy voiced stops might influence perceived duration of preceding vowels.

(a) There might be no effect of breathiness on perceived vowel duration, indicating that there are no acoustic effects of breathiness that influence perceived duration.

(b) Vowels before breathy voiced stops might sound longer than vowels before plain voiced stops, which would suggest that breathiness in the consonant or within the vowel increases perceived duration of the vowel.

(c) Vowels before breathy voiced stops might be perceived as shorter than vowels before plain voiced stops, which would suggest that breathiness in the consonant or within the vowel decreases perceived duration of the vowel.

2.4 Results

Table 2 presents the summary of a generalized linear mixed effects model for the 'long' responses to each item. The random effect was participant. The fixed effects were duration step; voicing of the coda (voiced, breathy voiced, voiceless, voiceless aspirated); vowel quality (/a, i, u/); and block number. The model uses treatment contrasts; the reference level for coda voicing was voiceless unaspirated and for vowel was /a/.

Table 2 ***glmer model for 'long' responses, Experiment 1. Reference levels for categorical factors: Coda = voiceless unaspirated, Vowel = /a/.***

	β	CI	SE	z-value	*p*
(Intercept)	-1.71	-2.43 – -1.28	0.2	-8.43	< 0.001***
DurationStep	0.3	0.25 – 0.4	0.011	26.2	< 0.001***
Coda-Voiced	-0.42	-0.61 – -0.27	0.086	-4.86	< 0.001***
Coda-VoicelessAsp	-0.15	-0.33 – 0.015	0.086	-1.72	0.085
Coda-BreathyVoiced	-0.0096	-0.19 – 0.15	0.086	-0.11	0.91
Vowel-i	0.65	0.53 – 0.83	0.075	8.71	< 0.001***
Vowel-u	0.86	0.74 – 1.04	0.075	11.4	< 0.001***
Block	-0.15	-0.23 – -0.079	0.037	-3.94	< 0.001***

While an interaction between the vowel quality and duration step for this is significant in a model that includes it, with larger effects of duration step for /a/ vowels, these interactions have very high correlations with the main factors for vowel quality and models with the interaction do not reliably converge, so the interaction has not been included.

Listeners were sensitive to the duration of the vowel. Responses were not clearly categorical as they are for listeners of languages which have phonological vowel length contrasts, but duration was a significant predictor of whether a vowel was identified as long or short; listeners were approximately 1.3 times more likely to give a ‘long’ response with each increasing step of duration. The proportion of ‘long’ responses along each step of the continuum is illustrated in Figure 1. The lack of categorical divide suggests that listeners are not mapping duration onto an existing English contrast.

Responses of ‘long’ were 0.66 times as likely for vowels before voiced stops as for vowels before voiceless stops. The pattern of decisions is consistent with listeners compensating for their expectations about the duration of vowels based on context; longer durations are expected before voiced stops, so listeners have a higher threshold for length in this environment. Effects of the coda are illustrated in Figure 1. Notably, this effect was apparent throughout most of the continuum, particularly at the intermediate durations; thus, it was not driven by the difference in starting durations for vowels in each environment, which caused the duration manipulations to differ for the longest two duration steps.

On the other hand, aspiration had no apparent effect on vowel duration. Voiceless aspirated stops patterned like voiceless unaspirated stops, with only marginally fewer ‘long’ responses. Breathy voiced stops also patterned with voiceless stops, rather than with voiced stops, which might suggest either that English-speaking listeners did not perceive these stops as belonging to the voiced category, or that characteristics of breathiness contribute to greater perceived duration and outweigh compensatory expectations based on voicing.

Vowel quality was also a predictor of responses. The high vowels /i/ and /u/, which are inherently shorter than the low vowel, were more frequently identified as long than the vowel /a/; ‘long’ responses were 1.9 times more likely for /i/ and 2.4 times more likely for /u/, which suggests that listeners are compensating for their expectations about the duration of each vowel.

The number of ‘long’ responses decreased across blocks. That is, listeners became less likely to respond ‘long’ later into the experiment.

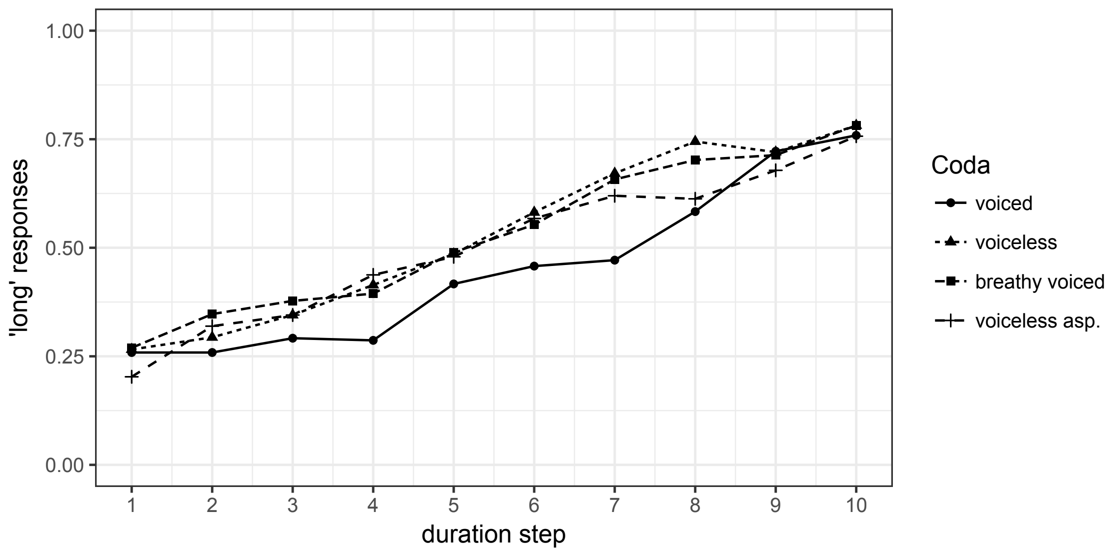

Figure 1. Proportion of ‘long’ responses in Experiment 1, by duration step and laryngeal features of the coda. Based on the raw data, not the output of the regression model; pooled across speakers.

3.0 Experiment 2

3.1 Recordings

These experiments used the same base recordings as in Experiment 1: a male Hindi speaker reading VC nonce words, elicited individually in randomized order and recorded in a sound attenuated booth with a stand-mounted Blue Yeti microphone in the Audacity software program and digitized at a 44.1 kHz sampling rate with 16-bit quantization.

The stimuli were made from the same recordings as in Experiment 1, but differed in manipulation of the codas. The vowels came from 3 coda environments: voiced /d/, voiceless /t/, and breathy voiced /d^{ɦ}/. The coda, including the closure and release burst was removed; the formant transition to the coda as also removed. The end of the steady state portion of the vowel was defined as point at which the frequency change in F1 was greater than 10 Hz from one 6.5 ms window to the next without returning to the earlier frequency within 5 windows, or where the frequency change in F2 was greater than 15 Hz per 6.5 ms without returning to the earlier frequency within 5 windows. The vowel portion of the stimulus was cut at this point and given a 50 ms fade-out. The point identified by these calculations and how that corresponds to the formant trajectories on a spectrogram are illustrated in Figure 2.

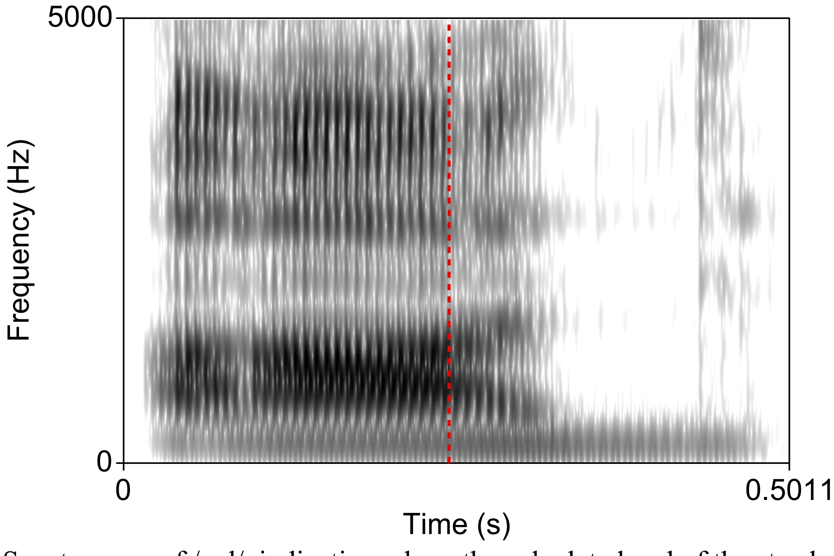

Figure 2. Spectrogram of /a:d/, indicating where the calculated end of the steady-state portion of the vowel; the vowel portion of the stimulus ended at this point.

Because the transition portions of the vowels were removed, the naturally produced long vowels were not long enough that all stimuli could be made by removing glottal cycles. Before voiced and breathy voiced codas, the naturally produced items aligned with duration step 9 for /a/ and aligned with step 8 for /i/ and /u/; before voiceless codas, the naturally produced items aligned with duration step 7 for /a/ and duration step 6 for /i/ and /u/. These differences in the manipulations based on vowel quality and coda voicing could potentially cause differences in perception, which should be apparent only at these long durations for which the manipulations differed.

Using the concatenate command in the software program Praat, each vowel was spliced with each of 3 different endings, designed to test Javkin's (1976) proposal that voicing-conditioned vowel duration differences are based on uncertainty about the divide between vowels and voiced coda consonants. While these spliced conditions do not provide an exact parallel for Javkin's proposal, as eliminating formant transitions within the vowels reduces cues for the coda, the lack of clear cues for the presence of a coda should maximize the possibility that listeners could interpret these endings as part of the vowel. Each ending was given the same duration of 70 ms; however, as the timing of trials was under control of the participant, rather than proceeding automatically, the uniform duration of the three endings is not likely to be crucial. The first ending contained 70 ms of the modal voicing from the closure portion of a coda /d/, with no release burst. No manipulation was made to enhance perception of this ending as a coda; indeed, the goal was to minimize the possibility that this ending would be perceived as a coda. The 70 ms duration was based on the mean duration of closures within the voiced stops within the speaker's original productions. The second ending was a 50 ms fade-in to the same voicing from the closure portion of a coda /d/, with 20 ms of stable voicing; this ending tested whether indistinctness of the boundary between vowels and following voiced consonants depends on continuous voicing or if the same effect is achieved if voicing ends and begins again. The third ending was silence; these items were presumably heard as ending with the vowel. The waveforms for each ending are given in Figure 3.

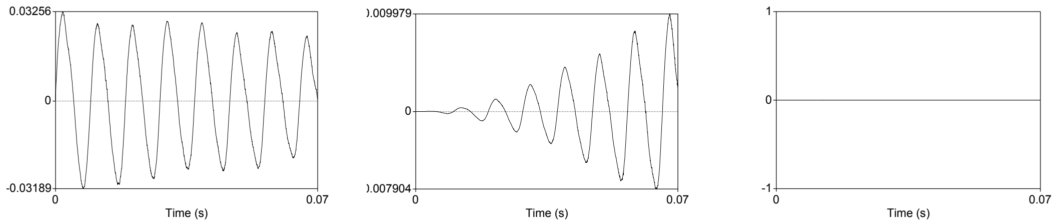

Figure 3. Waveforms for the three endings: (a) voicing from the closure of a coda /d/, (b) fade-in to the voicing from the closure of a coda /d/, (c) silence.

3.2 Experiment Design

For Experiment 2a, 24 native English speakers (8 male; mean age 22.3, age range: 18-36) participated in the study. None of them had any linguistic training. They completed a forced choice task about vowel duration, as in Experiment 1. Each participant completed two duration perception experiments; they were randomly assigned to two of the four experiments, and the order of the tasks was randomly determined.

The stimuli were made as described in the preceding section. There were three vowel qualities (/a/, /i/, /u/), and 3 coda environments in which the vowels had been produced: voiced /d/, voiceless /t/, and breathy voiced /d^{ɦ}/. The codas and the formant transitions to them were removed as described above. Each vowel was spliced with the three endings described above. The 10-step duration continuum was the same as in Experiment 1. This produced a total of 270 items, which were heard a single time by each listener. All other aspects of the procedure were the same as in Experiment 1.

For Experiment 2b, 24 native English speakers (9 male; mean age 23.5, age range: 18-41) participated in the study. None of them had any linguistic training. They completed a forced choice task about vowel duration, as in the other experiments. Each participant completed two duration perception experiments; they were randomly assigned to two of the four experiments, and the order of the tasks was randomly determined.

The same stimuli from Experiment 2a were used, excluding items from the breathy voiced /$d^{ɦ}$/ context, producing a total of 180 items for each listener. All other parameters of the stimuli and task were the same as in Experiment 1. Excluding the vowels from the breathy voiced environment was aimed to control for the possibility that the set of items present in the stimuli influenced what acoustic cues listeners were attending to; the presence of breathy vowels from the breathy voiced coda environment could attract listeners' attention to phonation cues present in the breathy vowels or otherwise influence their interaction with the stimuli.

Statistical results are based on logistic regression using a binomial distribution, calculated in R version 3.4.3 (R Core Team 2017) with the lme4 (version 1.1-7) package (Bates et al., 2015). *p*-values were calculated by the lmerTest (version 2.0-33) package (Kuznetsova et al., 2015). Responses with latencies shorter than 250 ms or longer than 5 s were excluded from analysis (< 1% of the data).

3.3 Hypotheses

Hypothesis 1: The voicing of the coda that was originally produced with the vowel might influence the perceived duration of the vowel, either based on listeners' phonological knowledge or perceptual biases due to the acoustic form.

(a) There might be no effect of the original voicing environment. This would indicate that none of the acoustic effects of coda voicing on aspects of the vowel influence perceived duration, and suggest that the voicing effect is either articulatorily driven or based on perceptual effects of the coda itself.

(b) Listeners might perceive the original coda as still being present, based on secondary cues, and compensate for expected duration in that environment, giving more 'long' responses to vowels produced before voiceless codas.

(c) If consonant voicing produces acoustic characteristics in preceding vowels that increases their perceived duration, then listeners should give more 'long' responses to vowels that were produced before voiced codas than to vowels produced before voiceless codas.

Hypothesis 2: Breathy voiced stops might influence perceived duration of preceding vowels differently than modally voiced stops.

(a) There might be no difference between the perception of vowels produced before voiced stops and produced before breathy voiced stops, indicating that there are no acoustic effects of breathiness that influence perceived duration.

(b) Vowels produced before breathy voiced stops might sound longer than vowels produced before plain voiced stops, which would suggest that breathiness within the vowel increases perceived duration of the vowel.

(c) Vowels from before breathy voiced stops might be perceived as shorter than vowels produced before plain voiced stops, which would suggest that breathiness within the vowel decreases perceived duration of the vowel.

Hypothesis 3: The spliced endings might influence perceived duration of the vowel, either if they induce listeners to perceive that codas are present or if they influence listeners' perception of where vowels end.

(a) There might be no effect of the spliced ending. Because the stimuli included no release burst or formant transitions to suggest the presence of a consonant, listeners might be entirely unaffected by the presence or absence of noise after the vowel. This would suggest that Javkin's (1976) proposal about indistinctness of segment boundaries driving voicing-conditioned duration effects is inaccurate.

(b) Listeners might perceive the spliced ending of the closure portion of /d/ as indicating the presence of a voiced coda, despite the lack of release burst or formant transitions, and compensate for it as they do for codas that are fully cued, giving fewer 'long' responses to items with the voiced ending.

(c) If Javkin's proposal is accurate, listeners should attribute the periodic noise in the voiced ending to the vowel, and give more 'long' responses to items with this ending. The lack of transitions or release burst to indicate the presence of a consonant that the periodic noise could be attributed to should maximize the possibility of this reanalysis.

Hypothesis 4: The range of productions present among the stimuli may influence results. The comparison between 2a, which contains vowels produced before breathy voiced codas, and 2b, which lacks these items, tested possible effects of the stimulus set. While some analyses can be made based on the comparison between these two experiments, they are speculative and would require further work to test; the main goal is to confirm that the presence of vowels from the breathy voiced environment is not driving a difference in perceived duration of vowels from the voiced and voiceless environments. There are three possible effects:

(a) The perception of vowels from the voiced and voiceless production environments could exhibit a consistent pattern relative to each other, regardless of the presence of vowels from the breathy voiced environment. This would suggest that listeners' attention to different acoustic details is not being influenced by variability in those characteristics across the stimuli.

(b) The difference in perceived duration of vowels from the voiced and voiceless environments could be present only when vowels from the breathy voiced environment are present. This would suggest that listeners' attention is being attracted to different acoustic details in each condition. In particular, they might be more influenced by spectral tilt, which is more prominently distinct in vowels from the breathy voiced environment than vowels in the other environments, and this would suggest that spectral tilt or other phonation cues contribute to a perceived difference between vowels from the voiced and voiceless environment.

(c) The difference in perceived duration of vowels from the voiced and voiceless environments could be present only when vowels from the breathy voiced environment are not present. This would also suggest that listeners attention is being attracted to different acoustic details in each condition. While the same details, and in particular spectral tilt, are likely to be more prominent when vowels from the breathy voiced environment are present, this effect would suggest that other cues are responsible for the perceived difference between vowels from the voiced and voiceless environment, and listeners' attention those cues is decreased when breathy vowels are present.

3.4 Results: Experiment 2a

Table 3 presents the summary of a generalized linear mixed effects model for 'long' responses to each item. The random effect was participant. The fixed effects were duration step; voicing of the original coda (voiced, breathy voiced, voiceless); voicing of the spliced-in ending (voicing, silence, delayed voicing); vowel quality (/a, i, u/); and block number. The model uses treatment contrasts; the reference level for coda voicing was voiced, for the spliced ending was voicing, and for vowel was /a/.

Table 3 ***glmer model for 'long' responses, Experiment 2a. Reference levels for categorical factors: OrigCoda = voiced, SplicedEnd = voicing, Vowel = /a/.***

	β	CI	SE	z-value	*p*
(Intercept)	-3.07	-3.42 – -2.73	0.18	-17.49	< 0.001***
DurationStep	0.48	0.46 – 0.51	0.013	38.1	< 0.001***
OrigCoda-Voiceless	-0.23	-0.37 – -0.081	0.075	-3.05	0.0023**
OrigCoda-BreathyVoiced	0.35	0.2 – 0.5	0.075	4.67	< 0.001***
SplicedEnd-DelayVoice	0.027	-0.12 – 0.17	0.075	0.36	0.72
SplicedEnd-Silence	-0.12	-0.27 – 0.024	0.075	-1.64	0.1
Vowel-i	0.22	0.075 – 0.37	0.075	2.97	0.003**
Vowel-u	-0.049	-0.2 – 0.097	0.075	-0.66	0.51
Block	0.24	0.16 – 0.31	0.038	6.31	< 0.001***

Even without the presence of codas to allow a comparative evaluation of vowel duration given the speech rate suggested by other segments, listeners were able to perceive vowel duration.

Listeners gave substantially more 'long' responses to items of greater duration; they were 1.6 times more likely to give a 'long' response with each increasing step of duration. The proportion of 'long' responses along each step of the continuum is illustrated in Figure 4.

Responses of 'long' were more likely for vowels produced before voiced stops than for vowels produced before voiceless stops; 'long' responses were 0.79 times as likely with vowels from the voiceless coda environment. The difference was most apparent in stimuli with moderate duration, for which the duration itself is most ambiguous; see Figure 4. This effect suggests that voicing produces acoustic characteristics that make vowels seem longer, when the coda has been removed and there is thus no conditioning environment to which the characteristics can be attributed. Notably, this is the opposite of the result in Experiment 1, in which the codas were still present. See Table 1 for a summary of the vowels' acoustic characteristics. The effect was not localized at the long duration steps, indicating that it was not driven by the difference in starting durations for vowels in each environment, which caused the duration manipulations to differ for some of the vowels for the longest four duration steps.

Breathy voicing of the original coda produced an even greater proportion of 'long' responses than plain voicing; 'long' responses were 1.4 times more likely for vowels from the breathy voiced environment. This is consistent with breathiness increasing perceived duration in vowels.

The spliced ending only had a marginal effect. There were slightly fewer 'long' identifications of vowels followed by silence than vowels followed by voicing from the closed portion of a stop; 'long' responses were approximately 0.89 times as likely in this environment.

Vowel quality was a less consistent predictor of responses in this experiment; only the high front /i/ elicited significantly more 'long' responses than /a/ (1.25 times more likely). The absence of a consonantal environment to establish relative duration may have weakened the role of listeners' expectations about inherent vowel duration. Alternatively, the lack of consonants may elicit responses that are less strongly influenced by phonological expectations, as the forms are less word-like.

The number of 'long' responses increased across blocks.

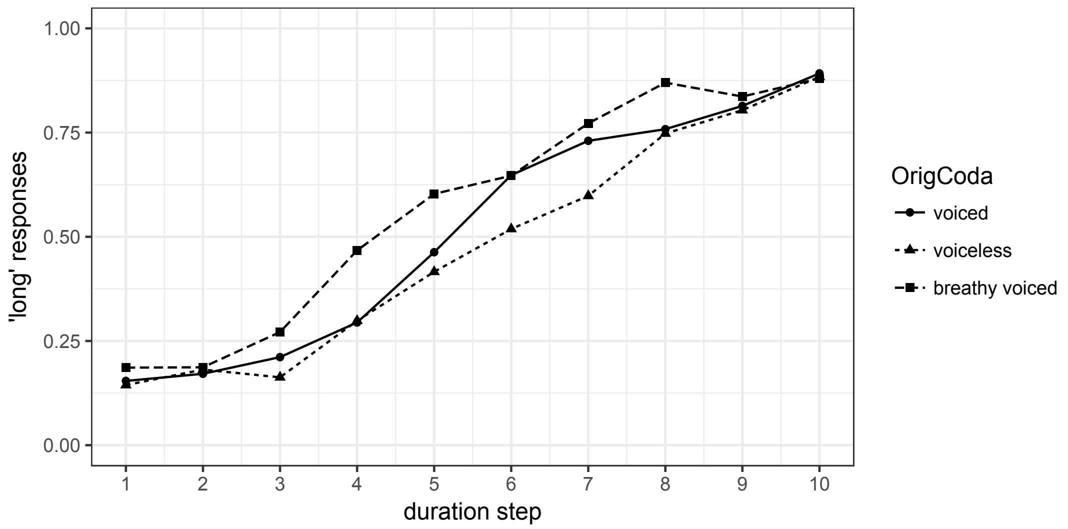

Figure 4. Proportion of 'long' responses in Experiment 2a, by duration step and voicing of the original coda. Based on the raw data, not the output of the regression model; pooled across speakers.

3.5 Results: Experiment 2b

Table 4 presents the summary of a generalized linear mixed effects model for 'long' responses to each item. The random effect was participant. The fixed effects were duration step; voicing of the original coda (voiced, voiceless); voicing of the spliced-in ending (voicing, silence, delayed voicing); vowel quality (/a, i, u/); and block number. The model uses treatment contrasts; the reference level for coda voicing was voiced, for the spliced ending was voicing, and for vowel was /a/.

Table 4 ***glmer model for 'long' responses, Experiment 2b. Reference levels for categorical factors: OrigCoda = voiced, SplicedEnd = voicing, Vowel = /a/.***

	β	CI	SE	z-value	*p*
(Intercept)	-2.11	-2.47 – -1.76	0.18	-11.6	< 0.001***
DurationStep	0.46	0.43 – 0.49	0.015	30.6	< 0.001***
OrigCoda-Voiceless	-0.51	-0.65 – -0.36	0.074	-6.88	< 0.001***
SplicedEnd-DelayVoice	-0.063	-0.24 – 0.11	0.090	-0.7	0.48
SplicedEnd-Silence	-0.26	-0.44 – -0.082	0.090	-2.88	0.0041**
Vowel-i	0.039	-0.14 – 0.22	0.090	0.43	0.67
Vowel-u	-0.21	-0.38 – -0.029	0.090	-2.26	0.024*
Block	-0.0077	-0.096 – 0.081	0.045	-0.17	0.86

Again, the actual duration was a strong predictor of the percentage of 'long' responses. Listeners were 1.6 times more likely to give a 'long' response with each increasing step of duration. The proportion of 'long' responses along each step of the continuum is illustrated in Figure 5.

The voicing of the original coda that had been produced with the vowel was a significant predictor of responses. There were substantially more 'long' responses to vowels produced before voiced stops than to vowels produced before voiceless stops; 'long' responses were 0.6 times as likely before voiceless stops. In contrast to Experiment 2a, the effect is apparent throughout the duration continuum, as is illustrated in Figure 5.

The voicing of the ending spliced onto the vowel was also a significant predictor of responses. There were more 'long' responses to vowels followed by voicing from the closed portion of a voiced stop than to vowels followed by silence; 'long' responses were 0.77 times as likely before silence. There was no difference between immediate voicing and voicing with a gradual fade-in.

Vowel quality was a weak predictor of responses in this experiment, as in Experiment 2a, though the division differed; there were somewhat fewer 'long' responses to /u/ than to /a/, while responses to /i/ and /a/ did not differ significantly.

The number of 'long' responses did not change across blocks.

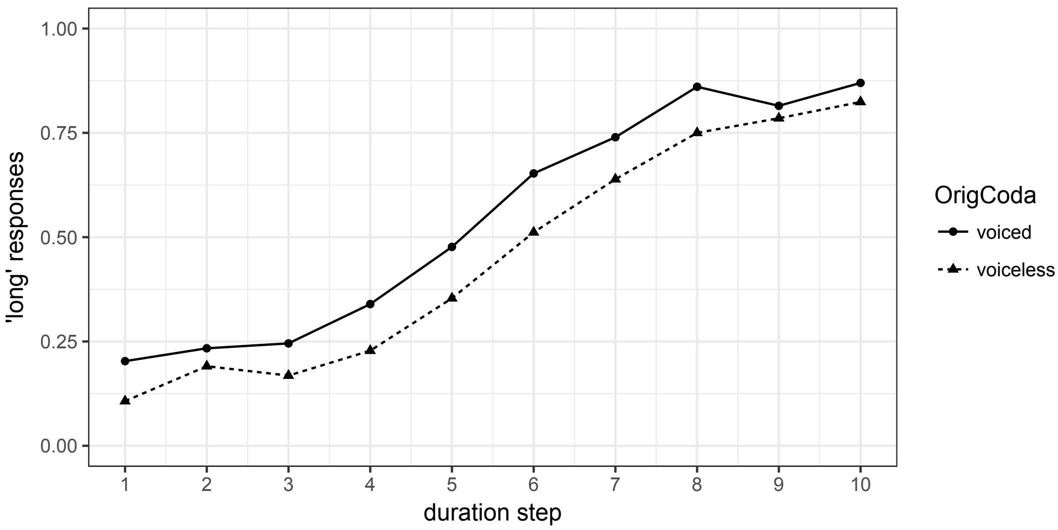

Figure 5. Proportion of ‘long’ responses in Experiment 2b, by duration step and voicing of the original coda. Based on the raw data, not the output of the regression model; pooled across speakers.

4.0 Experiment 3

4.1 Recordings

This experiment used the same base recordings used in the previous experiments: a male Hindi speaker reading VC nonce words, elicited individually in randomized order and recorded in a sound attenuated booth with a stand-mounted Blue Yeti microphone in the Audacity software program and digitized at a 44.1 kHz sampling rate with 16-bit quantization.

4.2 Experiment Design

24 native English speakers (9 male; mean age 24.0, age range: 18-41) participated in the study. None of them had any linguistic training. They completed a forced choice task about vowel duration, as in the other experiments. Each participant completed two duration perception experiment; they were randomly assigned to two of the four experiments, and the order of the tasks was randomly determined.

The same coda environments from Experiment 2b were used: vowels which had been produced before voiced and voiceless codas, with the codas themselves removed. The f0 was manipulated to create either a 30 Hz fall, a 30 Hz rise, or a nearly level f0. As flat f0 can make stimuli sound unnatural, items in the ‘level’ category had a 3 Hz f0 drop rather than being truly flat (cf. Kohler, 1985; van Dommelen, 1993); using manipulations of natural speech rather than complete synthesis also reduces the risk of stimuli not registering as speech.

The f0 contours were comparable in size to the f0 drop that occurs at English phrase boundaries, and are larger than the f0 contour difference between voiced and voiceless codas among these stimuli (as summarized in Table 1). The larger contour was used to increase the likelihood of observing an effect of f0 contour on perceived duration. As previous work has demonstrated effects of f0 contour on perceived duration, the main question of interest was to address how the f0 contour would interact with the effect of the original coda voicing. If the f0 differences among the stimuli were driving the effect of original coda in Studies 2a and 2b, these effects would be eliminated by the f0 contour manipulation. An interaction could also reflect a shift in how listeners are approaching the task; presenting them with f0 patterns that align with English intonational patterns might focus their attention on cues that they expect to align with intonational contrasts, rather than characteristics that they associate with segmental contrasts. The f0 rise is most similar to the realization of English stress, though it doesn't match as closely as the steady 30 Hz fall does to the phrase-final f0 drop.

Each vowel was spliced with 2 different endings: 70 ms of voicing and 70 ms of silence. The ending with a gradual voicing fade-in was excluded to keep the number of trials comparable to the other experiments. This produced a total of 360 items which each listener heard. All other parameters of the stimuli and task were the same as in Experiment 1.

Statistical results are based on logistic regression using a binomial distribution, calculated in R version 3.4.3 (R Core Team 2017) with the lme4 (version 1.1-7) package (Bates et al., 2015). *p*-values were calculated by the lmerTest (version 2.0-33) package (Kuznetsova et al., 2015). Responses with latencies shorter than 250 ms or longer than 5 s were excluded from analysis (< 1% of the data).

4.3 Hypotheses

Hypothesis 1: As in the previous experiments, the voicing of the coda that was originally produced with the vowel might influence the perceived duration of the vowel.

(a) There might be no effect of the original voicing environment, indicating that none of the acoustic effects of coda voicing on aspects of the vowel influence perceived duration.

(b) Listeners might perceive the original coda as still being present, based on secondary cues, and compensate for expected duration in that environment, giving more 'long' responses to vowels produced before voiceless codas.

(c) If consonant voicing produces acoustic characteristics in preceding vowels that increases perceived duration of those vowels, then listeners should give more 'long' responses to vowels that were produced before voiced codas.

Hypothesis 2: The spliced endings might influence perceived duration of the vowel, either if they induce listeners to perceive that codas are present or if they influence listeners' perception of where vowels end.

(a) There might be no effect of the spliced ending. Because the stimuli included no release burst or formant transitions to suggest the presence of a consonant, listeners might be entirely unaffected by the presence or absence of noise after the vowel. This would indicate that uncertainty about segment boundaries is not influencing perceived vowel duration.

(b) Listeners might perceive the spliced ending of the closure portion of /d/ as indicating the presence of a voiced coda, and compensate for the duration expected in that environment, giving fewer 'long' responses to items with the voiced ending.

(c) If Javkin's (1976) proposal is accurate, listeners should attribute the periodic noise in the voiced ending to the vowel, and give more 'long' responses to items with this ending.

Hypothesis 3: The f0 contour might influence listeners' perception of the duration of vowels, as has been observed in prior work.

(a) Listeners may give more 'long' responses to items with rising or falling contours than to items with a level f0, indicating that dynamics in f0 increase perceived vowel duration.

(b) Listeners may give more 'long' responses to items with falling contours than with rising contours, indicating that the particular f0 contour is crucial for how it influences perceived duration. The 30 Hz drop is consistent with English phrase-final intonation, which may also be associated with the longer duration typical in this environment.

(c) Listeners may give more 'long' responses to items with rising contours than with falling contours, indicating that the particular f0 contour is crucial for how it influences perceived duration. The rising f0 is the most similar to the contour found in stressed syllables; particularly if listeners are mapping duration onto stress, this could increase 'long' responses to rising f0 stimuli.

(d) The effect of original coda voicing on perceived vowel duration may interact with the f0 contour. As one of the major differences among these stimuli between vowels produced before voiced codas and before voiceless codas was the larger f0 fall before voiced codas, the f0 manipulation may eliminate the perceptual differences between the voicing environments. Such an interaction might also reflect the f0 manipulation obscuring other acoustic characteristics or directing attention away from them.

4.4 Results

Table 5 presents the summary of a generalized linear mixed effects model for 'long' responses to each item. The random effect was participant. The fixed effects were duration step; voicing of the original coda (voiced, voiceless); voicing of the spliced-in ending (voiced, voiceless); pitch (flat, falling, rising); vowel quality (/a, i, u/); and block number. The model uses treatment contrasts; the reference level for coda voicing was voiced, for the spliced ending was voicing, for pitch was flat, and for vowel was /a/.

Table 5 ***glmer model for 'long' responses, Experiment 2b. Reference levels for categorical factors: OrigCoda = voiced, SplicedEnd = voicing, Pitch = flat, Vowel = /a/.***

	β	CI	SE	z-value	*p*
(Intercept)	-3.11	-3.39 – -2.83	0.14	-21.6	< 0.001***
DurationStep	0.52	0.5 – 0.54	0.011	46.1	< 0.001***
OrigCoda-Voiceless	0.066	-0.039 – 0.17	0.053	1.23	0.22
SplicedEnd-Silence	-0.061	-0.17 – 0.044	0.053	-1.14	0.25
Pitch-Fall	0.22	-0.093 – 0.35	0.065	3.38	< 0.001***
Pitch-Rise	-0.024	-0.15 – 0.1	0.065	-0.37	0.71
Vowel-i	-0.018	-0.15 – 0.11	0.065	-0.28	0.78
Vowel-u	-0.074	-0.2 – 0.054	0.065	-1.13	0.26
Block	0.079	0.015 – 0.14	0.033	2.43	0.015*

The actual duration was a strong predictor of the percentage of 'long' responses. Listeners were 1.7 times more likely to give a 'long' response with each increasing step of duration. The proportion of 'long' responses along each step of the continuum is illustrated in Figures 6 and 7.

In contrast to the previous studies, the voicing of the original coda that had been produced with the vowel was not a significant predictor of responses. No clear difference was apparent at any point in the duration continuum, as is illustrated in Figure 6.

The voicing of the ending spliced onto the vowel was also not a significant predictor of responses.

However, the f0 contour was a significant predictor of responses. There were significantly more 'long' responses to vowels with a falling f0 than to vowels with a flat f0 (1.2 times more likely), while vowels with a rising f0 did not differ significantly from those with a flat f0. The effect of f0 contour is illustrated in Figure 7.

There were no significant differences in responses based on vowel quality.

The number of 'long' responses increased slightly across blocks.

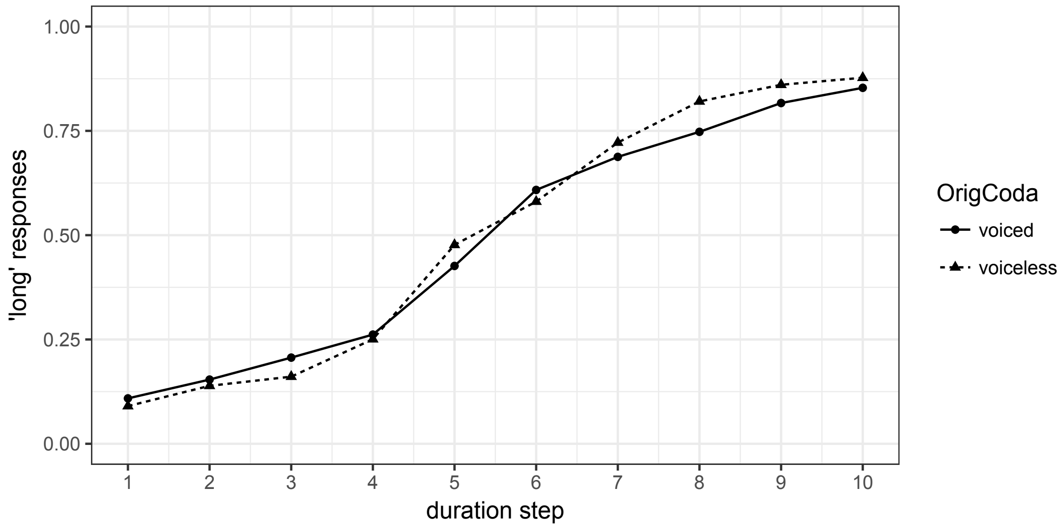

Figure 6. Proportion of ‘long’ responses in Experiment 3, by duration step and voicing of the original coda. Based on the raw data, not the output of the regression model; pooled across speakers.

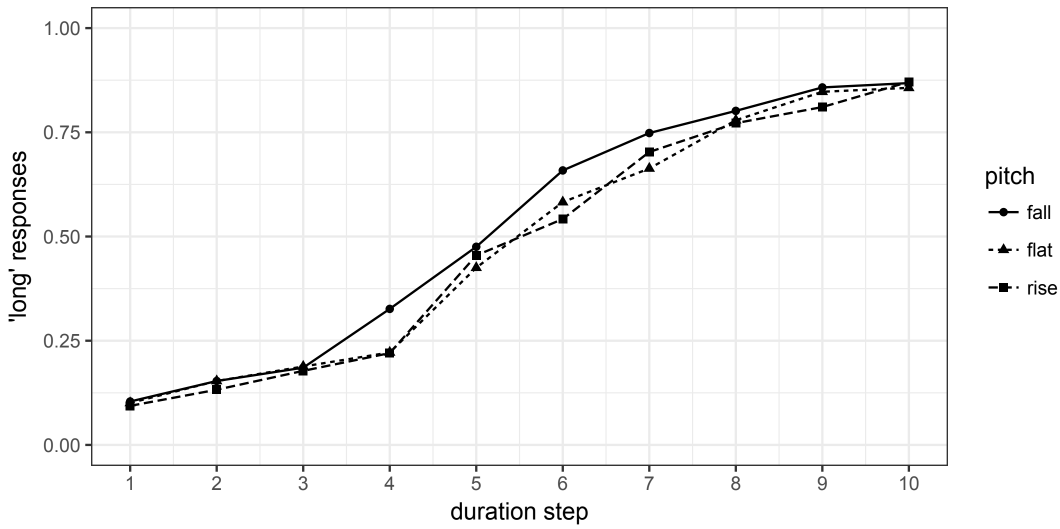

Figure 7. Proportion of ‘long’ responses in Experiment 3, by duration step and f0 manipulation. Based on the raw data, not the output of the regression model; pooled across speakers.

4.0 Discussion

The results of these experiments demonstrate several interacting factors that influence perceived vowel duration, based on expectations set by the coda itself as well as acoustic effects of the coda on the vowel, which are retained even when the coda is removed. The effects of the original coda on perceived vowel duration when the coda is removed suggest a possible perceptual pathway for the voicing effect on vowel duration.

English speakers can perceive vowel duration, as demonstrated by the large effect of actual duration on perceived duration of stimuli. They received no training aside from an explanation that they were listening purely to duration and not vowel quality. Sensitivity to duration is not dictated by phonology, though phonology can influence the shape of perception. Rather than a sharp categorical divide as found for speakers of languages with vowel length contrasts, English speakers have a largely linear continuum of duration identifications, which suggests that they are not directly mapping it onto stress or other English contrasts that involve duration.

Expectations based on knowledge of phonological patterns and typical phonetic details influence perception of duration. Listeners are sensitive to the typical durations for particular vowel qualities: /a/ is longer than high vowels, and accordingly an /a/ is less likely to be categorized as 'long' than an /i/ of equal length, though the effect is less clear for identifications of vowels in isolation. It is noteworthy that coda voicing is associated with lower F1 (higher vowels) within categories, which might also predict longer perceived duration of vowels before voiced codas; however, it remains for future work to test whether F1 differences within vowels have the same effects on perceived duration as F1 differences across vowels. If the differences in F1 based on coda voicing are a result of vowel duration rather than voicing itself (cf. Pycha & Dahan, 2016; Sanker, 2018), it is unlikely that F1 is a perceptual cue that results in longer duration before voiced codas.

When the consonantal context is present for listening, listeners compensate for it as well; they are sensitive to what the typical vowel duration in that environment is, so vowels are more likely to be categorized as 'long' before voiceless stops than before voiced stops. This parallels effects of vowel duration as used as a cue for coda voicing, with more identifications of codas as voiced when they are preceded by longer vowels (Raphael, 1972; Warner et al., 2004). This effect may also explain Fowler's (1992) result that longer closure durations increased perceived duration of vowels, if the manipulation of closure duration influenced perceived coda voicing, which listeners then compensated for. While listeners seem to compensate for expected vowel duration in each environment, the number of 'long' responses for vowels before voiced stops in Experiment 1 differed from the number for vowels before voiceless stops by only 1-2 steps along the duration continuum (~14-27 ms), which is much smaller than the typical differences in production of words in isolation; English vowels are 50-100 ms longer before voiced codas than before voiceless codas (Chen 1970; Peterson & Lehiste, 1960).

When the consonants and the transitions to them are removed, vowels are identified as longer when they came from an environment before a voiced stop than a voiceless stop, which suggests a possible perceptual source of conditioned vowel duration differences. If listeners interpret greater perceived vowel duration before voiced codas as reflecting the target duration, they could develop a voicing conditioned duration difference; such a difference could be further exaggerated if subsequent listeners do not fully compensate for differences in the environmentally expected duration. Such effects of course depend on acoustic differences existing in production; it doesn't exclude the possibility that there are also direct articulatory factors contributing to the voicing effect on duration. The acoustic characteristics of the stimuli suggest some possible factors that could be influencing perceived vowel duration: compared to vowels produced before voiceless codas, vowels produced before voiced codas have a larger drop in f0, a smaller drop in intensity, lower F1, and higher HNR. Any of these characteristics might contribute to differences in perceived duration.

In Experiment 3, in which the f0 contour was controlled, falling f0 was a predictor of more 'long' responses and the original voicing of the coda was eliminated as a predictor. This result could indicate that falling f0 within these stimuli is the primary way that the effects of the coda influenced perceived vowel duration, though it could also indicate that the variation in f0 contours attracted listeners' attention to f0 as a cue and decreased their attention to other cues that might have been driving the effect of the original coda on responses in Experiments 2a and 2b. In particular, the large f0 drop aligns with English phrase-final drops, which could influence listeners to expect the stimuli to align with intonational patterns. This could shift their focus away from other characteristics that they may attend to for segmental contrasts but that are not associated with intonational contrasts. The effects of other acoustic cues on perceived vowel duration, the interaction of those cues with coda voicing, and the extent to which listeners' attention to particular cues can be shifted be context, remain to be tested by future work.

The greater perceived duration of vowels only with falling f0, while vowels with rising f0 behaved like vowels with flat f0, is in contrast to previous studies that found both rising and falling contours to increase perceived duration (Yu, 2010; Lehiste, 1976), though some work has also found that falling f0 contours are perceived as longer than rising f0 contours (Cumming, 2011), which might suggest that the differences are dependent on the particular f0 contour. In this case, the falling f0 aligns with the English phrase-final contour; association of this contour with the longer durations that also occur in this environment may have increased how long listeners perceived these vowels as being.

Vowels also were perceived as longer before breathy voiced stops than other stops, whether or not the consonant is still present. The difference between effects of breathy voiced stops and plain voiced stops is consistent with the voicing effect on vowel duration being phonologized in English (Sharf, 1964; de Jong, 2004); listeners compensate for the predictable differences in duration based on following voiced vs. voiceless stops, while English-speaking listeners have no breathiness contrast, so they do not have any expectations about associated duration differences and do not compensate for effects of breathiness even when the coda is present. The acoustic characteristics provide some possible factors that could be influencing perceived duration: vowels produced before breathy voiced stops have a higher spectral tilt, higher jitter, and smaller intensity drop than vowels produced before modal voiced stops.

The presence of vowels from breathy voiced environments among the stimuli influences responses for other items; vowels from the plain voiced environment elicited fewer 'long' responses when the set of stimuli included vowels from the breathy voiced environment, perhaps because they were perceived as shorter in comparison, or because the presence of substantial breathiness in some stimuli directed listeners' attention to different acoustic cues than they otherwise were attending to. Vowels from the breathy voiced environment were more distinct from the other vowels in spectral tilt, with somewhat smaller differences in jitter and the intensity contour; the larger difference between the voiced and voiceless environments when vowels from the breathy voiced environment were not present might suggest that there were characteristics other than these that were contributing to difference in perceived duration between the voiced and voiceless environments. It is also possible that the long perceived duration of vowels from this environment made vowels from the modal voiced environment seem shorter in comparison, reducing the number of 'long' responses to the latter group.

Whether the vowel was followed by silence or by the closure portion of a /d/ had an inconsistent effect on perceived vowel duration. If listeners are prone to misanalysing the boundary between vowels and voiced codas (cf. Javkin, 1976), they should also be likely to interpret periodic noise following a vowel as being part of the vowel. In these stimuli, the formant transitions at the end of the vowels were removed and there was no release burst associated with the spliced ending, providing little evidence for a coda being present, which should maximize listeners' tendency to perceive periodic noise following the vowel as being part of the vowel.[3] Vowels followed by the ending with the closure portion of /d/ were more frequently identified as 'long' than vowels followed by silence in only one of the experiments. This result suggests that the voicing effect cannot be primarily attributed to listeners' uncertainty about where the boundary between vowels and following voiced consonants is, though it might make a small contribution. However, there was no difference between the ending with the closure portion of /d/ and the ending with a fade-in to the same voicing in any of the experiments; to the extent that there is an effect of periodic voicing noise following the vowel, the difference between silence and that voicing must depend on a longer silent period than was present in the stimuli with the voicing fade-in ending.

[3] One reviewer notes that a closer parallel for testing Javkin's (1976) hypothesis would be to use stimuli with normally cued codas, and test whether increased voicing duration within the coda closure increases perceived vowel duration when vowel duration is constant.

5.0 Conclusions

The results of these experiments suggest a possible perceptual pathway for the voicing effect on vowel duration. When the codas were removed, vowels that had been produced before a voiced coda were perceived as longer than vowels that had been produced before a voiceless coda, indicating that the acoustic characteristics caused by voiced codas, such as larger f0 drops, are not only salient, but influence perceived vowel length. If listeners interpret differences in perceived duration as reflecting differences in target duration, they could begin producing such a difference, phonologizing a voicing-conditioned vowel duration pattern.

The distinct pattern of responses when the codas were present suggests that English has phonologized vowel duration as conditioned by the voicing of following consonants. Thus, the coda sets expectations about vowel duration, and counteracts the effects of the acoustic details within the vowel on perceived vowel length. In contrast, vowels were perceived as longer when they were produced before breathy voiced stops, regardless of whether or not the stop was present in the stimulus, which is consistent with English speakers' lack of phonological rule for breathiness conditioning vowel length.

Acknowledgements

I am grateful to the audiences of the Annual Meeting on Phonology and the Linguistic Society of America and to three anonymous reviewers for their helpful comments and suggestions.

Appendix.

Figure 1. R command for the mixed effects model in Experiment 1.

```
glmer(winterperc1[which(winterperc1$responsetime > 0.25 &
winterperc1$responsetime < 5),]$Longresps ~ step + Coda + vowel +
block + (1|participant), data =
winterperc1[which(winterperc1$responsetime > 0.25 &
winterperc1$responsetime < 5),], family = binomial)
```

Figure 2. R command for the mixed effects model in Experiment 2a.

```
glmer(winterperc2a[which(winterperc2a$responsetime > 0.25 &
winterperc2a$responsetime < 5),]$Longresps ~ step + OrigCoda + ending
+ vowel + block + (1|participant), data =
winterperc2a[which(winterperc2a$responsetime > 0.25 &
winterperc2a$responsetime < 5),], family = binomial)
```

Figure 3. R command for the mixed effects model in Experiment 2b.

```
glmer(winterperc2b[which(winterperc2b$responsetime > 0.25 &
winterperc2b$responsetime < 5),]$Longresps ~ step + OrigCoda + ending
+ vowel + block + (1|participant), data =
winterperc2b[which(winterperc2b$responsetime > 0.25 &
winterperc2b$responsetime < 5),], family = binomial)
```

Figure 4. R command for the mixed effects model in Experiment 3.

```
glmer(winterperc3[which(winterperc3$responsetime > 0.25 &
winterperc3$responsetime < 5),]$Longresps ~ step + OrigCoda + ending +
pitch + vowel + block + (1|participant), data =
winterperc3[which(winterperc3$responsetime > 0.25 &
winterperc3$responsetime < 5),], family = binomial)
```

Reference List

Abel, S. M. (1972). Duration discrimination of noise and tone bursts. *Journal of the Acoustical Society of America*, 51, 1219–1223.

Abramson, A. S., & Ren, N. (1990). Distinctive vowel length: Duration versus spectrum in Thai. *Journal of Phonetics*, 18, 79–92.

Al-Tamimi, J., & Khattab, G. (2018). Acoustic correlates of the voicing contrast in Lebanese Arabic singleton and geminate stops. *Journal of Phonetics*, 71, 306–325.

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

Benkí, J. R. (2001). Place of articulation and first formant transition pattern both affect perception of voicing in English. *Journal of Phonetics*, 29, 1–22.

Best, C. (1995). A direct-realist view of cross-language perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Baltimore: York Press.

Chang, C. B. (2018). Perceptual attention as the locus of transfer to nonnative speech perception. *Journal of Phonetics*, 68, 85–102.

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

Choi, J., Kim, S., & Cho, T. (2016). Phonetic encoding of coda contrast under different focus conditions in L1 vs. L2 English. *Frontiers in Psychology*, 7, Article 624.

Cho, T., Whalen, D. H., & Docherty, G. (2019). Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics*, 72, 52–65.

Chong, A. J., Garellek, M. 2018. Online perception of glottalized coda stops in American English. *Laboratory Phonology*, 9(1), Article 4.

Coleman, J. 2003. Discovering the acoustic correlates of phonological contrasts. *Journal of Phonetics*, 31, 351–372.

Cooper, W. E. & Danly, M. (1981). Segmental and temporal aspects of utterance-final lengthening. *Phonetica*, 38, 106–115.

Cumming, R. (2011). The effect of dynamic fundamental frequency on the perception of duration. *Journal of Phonetics*, 39(3), 375–387.

Davidson, L. (2011). Phonetic, phonemic, and phonological factors in cross-linguistic discrimination of phonotactic contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 37(1), 270–282.

van Dommelen, W. A. (1993). Does dynamic f0 increase perceived duration? New light on an old issue. *Journal of Phonetics*, 21, 367–386.

Durvasula, K., & Luo, Q. (2014). Voicing, aspiration, and vowel duration in Hindi. *Proceedings of Meetings on Acoustics*, 18, 060009.

Fourakis, M., & Iverson, G. K. (1984). On the 'incomplete neutralization' of German final obstruents. *Phonetica*, 41, 140–149.

Fowler, C. A. (1992). Vowel duration and closure duration in voiced and unvoiced stops: There are no contrast effects here. *Journal of Phonetics*, 20, 143–165.

Gruenenfelder, T., & Pisoni, D. (1980). Fundamental frequency as a cue to postvocalic consonantal voicing: Some data from speech perception and production. *Perception and Psychophysics*, 28(6), 514–520.
Gussenhoven, C. (2004). Perceived vowel duration. In H. Quené & V. van Heuven (Eds.), *On Speech and Language: Studies for Sieb G. Nooteboom* (pp. 65–71). Utrecht: LOT.
Gussenhoven, C., & Zhou, W. (2013). Revisiting pitch slope and height effects on perceived duration. In F. Bimbot, C. Cerisara, C. Fougeron, G. Gravier, L. Lamel, F. Pellegrino, & P. Perrier (Eds.), *Proceedings of INTERSPEECH 2013: 14th Annual Conference of the International Speech Communication Association, Lyon, France, August 25-29, 2013* (pp. 1365–1369). ISCA.
Hadding-Koch, K. & Abrahamson, A. S. (1964). Duration versus spectrum in Swedish vowels: Some perceptual experiments. *Studia Linguistica*, 18, 94–107.
Halle, M., & Stevens, K. (1967). On the mechanism of glottal vibration for vowels and consonants. *MIT Research Laboratory of Electronics, Quarterly Progress Reports*, 85, 267–271.
Hombert, J.-M., Ohala, J., & Ewan, W. (1979). Phonetic explanations for the development of tones. *Language*, 55(1), 37–58.
Hussein, L. (1994). *Voicing-dependent Vowel Duration in Standard Arabic and its Acquisition by Adult American Students*. Ph.D. dissertation, Ohio State University.
Javkin, H. (1976). The perceptual basis of vowel duration differences associated with the voiced/voiceless distinction. *Report of the Berkeley Phonology Laboratory*, 1, 78–92.
de Jong, K. (1991). An articulatory study of consonant-induced vowel duration changes in English. *Phonetica*, 48, 1–17.
de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics*, 32, 493–516.
de Jong, K. & Zawaydeh, B. (2002). Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics*, 30, 53–75.
Keating, P. A. (1979). *A Phonetic Study of Voicing Contrast in Polish*. Ph.D. dissertation, Brown University.
Keating, P. (1985). Linguistic and nonlinguistic effects on the perception of vowel duration. *UCLA working papers in phonetics*, 60, 20-39.
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.
Kohler, K. J. (1985). F0 in the perception of lenis and fortis plosives. *Journal of the Acoustical Society of America*, 78(1), 21–32.
Kong, E. J., Beckman, M. E., & Edwards, J. (2012). Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of Phonetics*, 40, 725–744.
Kuznetsova, A., Bruun Brockhoff, P., and Haubo Bojesen Christensen, R. (2015). *lmerTest: Tests in Linear Mixed Effects Models*. R package version 2.0-29.
Laeufer, C. (1992). Patterns of voicing-conditioned vowel duration in French and English. *Journal of Phonetics*, 20, 411–440.
Lehiste, I. (1976). Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics*, 4(2), 113–117.
Lisker, L. (1986). 'Voicing' in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, 29(1), 2–11.
Maddieson, I., & Gandour, J. (1976). Vowel length before aspirated consonants. *UCLA Working Papers in Phonetics,* 31, 47–52.
Maddieson, I. (1977). Further studies on vowel length before aspirated consonants. *UCLA Working Papers in Phonetics*, 38, 82–90.
Machač, P. and Skarnitzl, R. (2007). Temporal compensation in Czech? In J. Trouvain & W. J. Barry (Eds.), *Proceedings of ICPhS XVI, Saarbrücken, Germany, August 6-10, 2007* (pp. 537–540).
Mitleb, F. (1984). Voicing effect on vowel duration is not an absolute universal. *Journal of Phonetics*, 12, 23–27.

Mohr, B. (1971). Intrinsic variation in the speech signal. *Phonetica*, 23, 65–93.
Moreton, E. (2004). Realization of the English postvocalic [voice] contrast in F1 and F2. *Journal of Phonetics*, 32, 1–33.
Ohala, M. & Ohala, J. (1992). Phonetic universals and Hindi segment duration. In J. J. Ohala, T. Nearey, B. Derwing, M. Hodge, & G. Wiebe (Eds.), *Proceedings of the International Conference on Spoken Language Processing, Banff, 12-16 October 1992* (pp. 831–834). Alberta: University of Alberta.
Nowak, P. (2006). *Vowel Reduction in Polish*. Ph.D. dissertation, University of California at Berkeley.
Penney, J., Cox, F., Miles, K., & Palethorpe, S. (2018). Glottalisation as a cue to coda consonant voicing in Australian English. *Journal of Phonetics*, 66, 161–184.
Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32(6), 693–703.
Pierce, J. W. (2007). PsychoPy – Psychophysics software in Python. *Journal of Neuroscience Methods*, 162, 8–13.
Port, R. F., Al-Ani, S., & Maeda, S. (1980). Temporal compensation and universal phonetics. *Phonetica*, 37, 235–252.
Pycha, A., & Dahan, D. (2016). Differences in coda voicing trigger changes in gestural timing: A test case from the American English diphthong /aɪ/. *Journal of Phonetics*, 56, 15–37.
R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rammsayer, T. H., & Lima, S. D. (1991). Duration discrimination of filled and empty auditory intervals: Cognitive and perceptual factors. *Perception & Psychophysics*, 50(6), 565–574.
Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America*, 51(4), 1296–1303.
Sanker, C. (2018). Effects of laryngeal features on vowel duration: Implications for Winter's Law. *Papers in Historical Phonology* 3, 180–205.
Seyfarth, S., & Garellek, M. (2018). Plosive voicing acoustics and voice quality in Yerevan Armenian. *Journal of Phonetics*, 71, 425–450.
Sharf, D. J. (1964). Vowel duration in normal and whispered speech. *Language and Speech,* 7, 89–97.
Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America*, 82(3), 847–863.
Summers, W. V. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *Journal of the Acoustical Society of America,* 82(3). 847–863.
Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic duration differences in production and perception: Evidence from Dutch. *Journal of Phonetics,* 32, 251–276.
Yu, A. C. L. 2010. Tonal effects on perceived vowel duration. In C. Fougeron, B. Kuehnert, M. Imperio, & N. Vallee (Eds.), *Laboratory Phonology 10* (pp. 151-168). New York: de Gruyter.