
JOURNAL ARTICLE

A perceptual pathway for voicing-conditioned vowel duration

Chelsea Sanker

Department of Linguistics, Yale University, New Haven, CT, US
chelsea.sanker@yale.edu

When codas and vowels are cross-spliced, vowels originally produced with voiced codas are perceived as longer than vowels of the same duration produced with voiceless codas. The spliced coda has the opposite effect: Vowels presented with voiced codas are perceived as shorter. To explain what characteristics make vowels produced with voiced codas sound longer than vowels produced with voiceless codas, four experiments tested how acoustic correlates of voicing affect English speakers' perception of vowel duration. Vowels were manipulated in a ten-step duration continuum, and listeners categorized each vowel as 'long' or 'short.' Study 1 tested effects of vowel height categories (/æ, ε, ɪ/) and within-category F1. Study 2 tested effects of intensity contour. Study 3 tested effects of spectral tilt. Perceived vowel duration increased with vowel height and rising intensity. Perceived vowel duration decreased with falling intensity and with higher spectral tilt. There was no effect of within-category F1. Study 4 confirmed that the effect of the original coda environment is not specific to English stimuli. The effects of spectral tilt and intensity contour on perceived duration provide a possible perceptual pathway for the development of voicing-conditioned vowel duration; coda voicing influences these characteristics and they in turn influence perceived vowel duration.

Keywords: Vowel duration; coda voicing; voicing-conditioned vowel duration; speech perception

1. Introduction

Vowels are longer before voiced codas than before voiceless codas in many languages, though there is variation by language in the size of the effect (e.g., Chen, 1970; Keating, 1979), how duration differences interact with stress (e.g., de Jong, 2004; de Jong & Zawaydeh, 2002) or intrinsic duration differences across vowel qualities (e.g., Peterson & Lehiste, 1960; Port, Al-Ani, & Maeda, 1980), and how vowel duration behaves in contexts where voicing contrasts are neutralized (e.g., Sharf, 1964; Warner, Jongman, Sereno, & Kempes, 2004). These language-specific patterns of voicing-conditioned vowel duration indicate that it cannot be entirely constrained by physical factors, and is likely to be part of the phonological representation of coda voicing in some of these languages. Nevertheless, the pervasiveness of voicing-conditioned vowel duration across languages indicates that there must be strong factors driving it. Explanations for voicing-conditioned vowel duration have been proposed based on articulation (e.g., Chen, 1970; Coretta, 2019) and perception (e.g., Javkin, 1976; Kluender, Diehl, & Wright, 1988), but the underlying cause remains in question. There may be several contributing factors.

In this paper, I argue that voicing-conditioned vowel duration is the result of coda voicing influencing characteristics of the vowel other than duration, which then influence the perceived duration of the vowel. If the perceived duration of the vowel in each environment is interpreted as the intended duration for that vowel, voicing-conditioned

vowel duration could thus become phonologized. Some of the acoustic cues that may be responsible for that effect are F1 (Bauer, 2011; Summers, 1987), vowel intensity (Hillenbrand, Ingrisano, Smith, & Flege, 1984), and spectral tilt (Chong & Garellek, 2018; Coleman, 2003). These particular characteristics are of interest because they not only are influenced by coda voicing but also are likely to influence perceived vowel duration, based on previous work across vowel categories for F1 (Gussenhoven, 2007; Wang, Lehiste, Chuang, & Darnovsky, 1976) and with non-linguistic stimuli for intensity (Berglund, Berglund, Ekman, & Frankehaeuser, 1969; Goldstone & Lhamon, 1974). Effects of intensity are also likely to produce effects of spectral tilt, because perceived intensity increases with frequency (Fletcher & Munson, 1933; Robinson & Dadson, 1956). The four experiments presented here are aimed at identifying specific acoustic characteristics that could be perceptual sources of voicing-conditioned vowel duration. In all four experiments, vowels produced with voiced codas sound longer than vowels produced with voiceless codas, both with English stimuli and with Telugu stimuli. The existence of this perceptual effect with Telugu stimuli demonstrates that it is not merely a consequence of English having voicing-conditioned vowel duration in production; Telugu lacks a substantial voicing-conditioned vowel duration difference (Reddy, 1988; Sanker, 2018).

1.1. The status of voicing-conditioned vowel duration

Many languages exhibit large voicing-conditioned vowel duration differences, e.g., English (Chen, 1970; Peterson & Lehiste, 1960), Dutch (Warner et al., 2004), and Japanese (Port et al., 1980). Other languages have a small or negligible difference in vowel duration based on coda voicing, e.g., Polish and Czech (Keating, 1979), Arabic (Mitleb, 1984), and Telugu (Reddy, 1988). Some of the differences in results may be due to experimental design rather than differences across languages (Laeufer, 1992), but the size of the effect certainly differs across languages, which suggests that the forces driving it are not absolute.

How the size of voicing-conditioned vowel duration interacts with other characteristics can also indicate phonological status; phonologized duration differences can target a particular duration ratio, while an automatic process is likely to produce a consistent absolute difference in duration (Solé, 2007). The voicing effect in English is preserved with a consistent duration ratio across vowels of different inherent durations (House, 1961; Peterson & Lehiste, 1960), across different stress contexts (de Jong, 2004), and across positions within the phrase (Cooper & Danly, 1981). In contrast, the small difference in Arabic does not scale proportionally in vowels with different phonological length (Port et al., 1980) or differences in phrasal stress (de Jong & Zawaydeh, 2002). A lack of proportional scaling has also been observed in languages with a larger voicing effect, e.g., Dutch (Warner et al., 2004).

Language-specific variation in how voicing-conditioned vowel duration interacts with voicing neutralization also suggests that it is phonologized in some languages but not in others. For example, vowel duration differences in English are preserved in whispered speech, even though the realization of voicing contrasts has been altered (Sharf, 1964). German preserves vowel duration differences before word-final obstruents, despite final devoicing (Fourakis & Iverson, 1984); however, voicing contrasts within the consonants themselves are not fully neutralized either (Port & O'Dell, 1985), which could contribute to preserving vowel duration differences. In contrast, Dutch almost entirely neutralizes vowel duration differences before devoiced word-final obstruents (Warner et al., 2004). French similarly neutralizes vowel duration differences preceding consonants that have undergone voicing assimilation (Abdelli-Beruh, 2003).

Perceptual compensation for expected duration can reflect listeners' knowledge of a consistent relationship in production. Such compensation is seen based on vowel height;

low vowels are longer than higher vowels, which makes listeners less likely to identify them as long (Gussenhoven, 2007; Wang et al., 1976). Similarly, listeners seem to compensate for expected vowel duration in a given coda environment. Because longer vowels are expected before voiced codas, listeners are more likely to identify vowels as short in this environment (Sanker, 2019b); however, that study used vowels with their naturally produced codas, which leaves some uncertainty about the respective contributions of the voicing of the coda presented with the vowel and the coda originally produced with the vowel.

1.2. Possible sources of voicing-conditioned vowel duration

Various explanations for voicing-conditioned vowel duration have been proposed, but no explanation has been definitively confirmed. Some proposed explanations make predictions that are not borne out by additional data; some remain plausible but make predictions that have not yet been tested. It is possible that multiple factors contribute to the existence of voicing-conditioned vowel duration, as has been pointed out by Coretta (2019), among others.

One proposed explanation for voicing-conditioned vowel duration is that it results from differences in the coordinative timing of vowels with voiced and voiceless codas (e.g., Pycha & Dahan, 2016). Sometimes the differences are explained as a result of compensatory timing, with longer vowels before shorter consonants, as voiced obstruents have shorter constriction than voiceless obstruents (e.g., Coretta, 2019; Fowler, 1981). However, duration of vowels relative to following consonants is not consistent when considering other consonant characteristics: Vowels are longer before fricatives than before stops, though many fricatives are longer than stops (Umeda, 1975), and vowels are longer before voiced aspirated stops than before plain voiced stops (Durvasula & Luo, 2014). Transitions from vowels into voiced codas take longer than transitions into voiceless codas (de Jong, 1991; Summers, 1987), which provides some support for the hypothesis that vowel duration differences are caused at least in part by differences in relative timing. These differences in timing have been explained as the result of the transition from vowel voicing to obstruent voicing requiring more precise adjustment in the state of the larynx than the transition from vowel voicing to voicelessness does (Chomsky & Halle, 1968; Halle & Stevens, 1967) or greater force required to maintain the consonantal constriction for voiceless obstruents than for voiced obstruents (Chen, 1970; Öhman, 1967). However, much of the work on possible articulatory causes of duration differences comes from English, so it can be difficult to establish a causal relationship between vowel duration, transition duration, and coda voicing.

Kluender et al. (1988) suggest a perceptual explanation: Because voiced obstruents have shorter constrictions than voiceless obstruents, vowels might sound longer relative to these shorter consonants. However, Fowler (1992) found that longer stop closures increased listeners' perceived duration of preceding vowels, rather than decreasing it. Sanker's (2019b) finding that vowels are more likely to be perceived as long when they have a voiceless coda may help account for this result; longer closures increase listeners' perception that the coda is voiceless (Port & Dalby, 1982), and listeners compensate for the duration expected in the given coda environment.

Javkin (1976) proposes that voicing-conditioned vowel duration arises from misperception of the boundary between vowels and following voiced consonants, with periodic noise in the coda interpreted as part of the vowel. Such misperception would predict that vowels should be longer before sonorants, which are acoustically more similar to vowels, than before obstruents; however, such a difference in duration is not observed (House & Fairbanks, 1953; Umeda, 1975). If listeners misperceive coda voicing as part of

the vowel, then adding periodic noise to the end of a vowel without adding any other cues for the presence of a coda should increase the perceived duration of the vowel; however, Sanker (2019b) finds an inconsistent effect of such spliced endings on perceived vowel duration.

Another possibility is that vowel duration is influenced by coda voicing less directly. Coda voicing is correlated with a range of vowel characteristics other than vowel duration, which will be laid out in the following section. Several of these vowel characteristics have been linked to perceived duration, which could then produce a relationship between coda voicing and vowel duration. If coda voicing influences acoustic characteristics within the preceding vowel and these characteristics influence the perceived duration of vowels, vowels produced before voiced codas may sound longer than vowels produced before voiceless codas, as was found by Sanker (2019b). If listeners then interpret these distinct perceived durations in each environment as reflecting the intended duration, they could develop a phonological representation in which vowels are longer before voiced codas than before voiceless codas.

1.3. Other acoustic correlates of voicing

Coda voicing is correlated with more vowel characteristics than just duration. Several of these characteristics provide likely candidates for influences on perceived vowel duration, which might contribute to the development of voicing-conditioned vowel duration, as discussed in the preceding section.

Coda voicing influences F1. At least in low monophthongs, F1 is higher before voiceless consonants than before voiced consonants (Summers, 1987), though diphthongs ending in high vowels can exhibit the opposite effect (Moreton, 2004) and high or high-mid monophthongs sometimes exhibit no clear difference based on voicing environment (Bauer, 2011). F1 also influences perception of coda voicing; listeners are more likely to perceive a coda consonant as voiced when the F1 of the preceding vowel is lower (Benkí, 2001; Summers, 1988), though this effect can be absent for high vowels (Hillenbrand et al., 1984) and it varies somewhat by language (Crowther & Mann, 1992). Because effects of coda voicing are often measured within English, some of the apparent effects of voicing may actually be effects of duration. Pycha and Dahan (2016) demonstrate that F1 is affected similarly by coda voicing and by other durational factors such as speech rate and phrasal position.

F1 is also related to vowel duration. Higher vowels—which have a lower F1—are shorter than lower vowels (House & Fairbanks, 1953; Solé & Ohala, 2010). Across phonologically distinct vowel categories, high vowels are perceived as longer than lower vowels, consistent with listeners compensating for their knowledge of the relationship in production (Gussenhoven, 2007; Wang et al., 1976). There is less evidence for whether or not there is a relationship between F1 and vowel duration within categories. Existing work suggests that there is no correlation between production of F1 and duration within vowel categories; that is, by-token F1 within a vowel category does not correlate with duration (Toivonen et al., 2015). No work to date has established whether or not there is a correlation in perception.

Coda voicing influences intensity, likely because the inhibition of voicing begins before the end of the vowel. Vowels before voiceless codas have a steeper drop in intensity, ending with a lower final intensity than vowels before voiced codas (Archer, Zamuner, Engel, Fais, & Curtin, 2016; House & Fairbanks, 1953). There is also some evidence for use of intensity as a perceptual cue for coda voicing: Hillenbrand et al. (1984) found that intensity decay time was a predictor of decisions about coda voicing, with more rapid intensity drops increasing listeners' perception that codas were voiceless.

Rising intensity increases perceived duration and falling intensity decreases perceived duration, at least for non-linguistic stimuli (Grassi & Darwin, 2006; Schlauch, Ries, & DiGiovanni, 2001). Perceived duration is also influenced by mean intensity; louder stimuli sound longer than quieter stimuli (Berglund et al., 1969; Goldstone & Lhamon, 1974). Even if consonantal environment influences mean intensity, normalization of mean intensity across items during stimulus preparation in many experiments could eliminate perceptual differences. Normalizing mean intensity could also confound effects of intensity contour, as items with level intensity will end up with lower peak intensity than items with more dynamic intensity.

Spectral tilt (H1-H2) is lower before glottalized voiceless codas than before voiced codas, in addition to several other correlates of glottalization (Chong & Garellek, 2018; Seyfarth & Garellek, 2018), and glottalization decreases listeners' perception that codas are voiced (Chong & Garellek, 2018; Penney, Cox, & Szakay, 2018). However, spectral tilt is higher before non-glottalized voiceless codas than before voiced codas (Coleman, 2003; Sanker, 2019a). Higher spectral tilt is also associated with voiceless onsets, even when also controlling for F0 as a factor; the extent of the relationship varies by language (Kong, Beckman, & Edwards, 2012).

Spectral tilt might also exhibit an influence on perceived duration, as a result of the relationship between frequency and perceived intensity. Perceived intensity increases with frequency (Fletcher & Munson, 1933; Robinson & Dadson, 1956), so vowels with lower spectral tilt are likely to sound louder than vowels with high spectral tilt. Auditory stimuli with greater intensity are perceived as having longer duration (Berglund et al., 1969; Goldstone & Lhamon, 1974), so vowels with greater perceived loudness are likely to also be perceived as longer.

Much of the work on these correlates of coda voicing focuses on English, which poses a risk of conflating vowel characteristics that are the result of duration and characteristics that are the result of coda voicing. However, the existence of languages with little influence of coda voicing on vowel duration makes it possible to distinguish acoustic effects of coda voicing on vowel characteristics from acoustic effects of vowel duration (Sanker 2018: 194–198). Telugu stimuli are used in Experiment 4, in order to separate effects of the voicing of the original coda from effects of the original duration of the vowel; even when codas are removed and vowel duration is manipulated into equal duration steps, the original coda and the original duration may nonetheless influence acoustic characteristics of the vowels. In Telugu, coda voicing has only a small effect on duration of preceding vowels, while there is a large duration difference between phonologically long and short vowels. Reddy (1988) finds that the mean duration of long vowels before voiceless stops is 300 ms and before voiced stops is 320 ms, and the mean duration of short vowels before voiceless stops is 80 ms and before voiced stops is 85 ms. Sanker (2018) finds a similarly small effect: The mean duration of long vowels before voiceless unaspirated stops is 288 ms and before voiced unaspirated stops is 307 ms, and the mean duration of short vowels before voiceless unaspirated stops is 102 ms and before voiced unaspirated stops is 117 ms.

1.4. This set of studies

This article examines the possibility that voicing-conditioned vowel duration could be produced via a perceptual pathway: Coda voicing influences several acoustic characteristics of vowels, which might then influence the perceived duration of those vowels and result in reanalysis of vowel duration as part of the representation of coda voicing.

The four duration perception experiments presented here are aimed at testing several acoustic characteristics that are likely to influence perceived vowel duration, based on

correlations in production or perceptual effects with non-linguistic stimuli. As these characteristics are also related to coda voicing, their effects have implications for the indirect effects of coda voicing on perceived vowel duration. Direct effects of coda voicing on perceived vowel duration are also examined, both voicing of the original coda produced with the vowel and voicing of the spliced coda presented with the vowel. In the first three experiments, the effects of the voicing of the original coda are tested with English stimuli. The fourth experiment uses Telugu stimuli, to test whether the effect of the original coda is caused by effects of coda voicing itself on preceding vowels or if it is the result of voicing-conditioned vowel duration in English; in Telugu, vowel duration exhibits only a small effect of coda voicing (Reddy, 1988; Sanker, 2018), which makes it possible to separate effects of coda voicing from effects of voicing-conditioned vowel duration. All experiments use native English-speaking listeners, so the results demonstrate factors in English speaker's perception of vowel duration; while some of the non-linguistic parallels suggest that similar patterns will be observed for speakers of other languages, it remains for future work to examine whether the perceptual effects are the same or not across languages.

2. Methods

The study consists of four experiments, all following the same basic design. Each one tests a different aspect of how listeners perceive vowel duration.

Experiment 1 tests the influences of F1 and vowel quality on perceived vowel duration, as well as the influence of the original coda produced with the vowel. F1 is lower before voiced codas than before voiceless codas (Bauer, 2011; Summers, 1987). As a correlate of coda voicing, F1 is a potential source of differences in perceived vowel duration based on coda voicing. Such a possibility is further supported by the relationship between F1 and vowel duration. Vowel height across vowel categories is correlated with vowel duration in production (e.g., House & Fairbanks, 1953; Solé & Ohala, 2010), and also has been demonstrated to influence perceived duration (Gussenhoven, 2007; Wang et al., 1976). However, previous work has not tested whether F1 within a vowel category influences perceived vowel duration; Experiment 1 provides this test, using variable F1 in three vowel qualities (/æ, ε, ɪ/).

Experiment 2 tests the influence of intensity contour on perceived vowel duration, as well as the influence of the original coda produced with the vowel and the effect of voiced and voiceless spliced endings. Vowels produced before voiceless codas have a steeper drop in intensity and a lower final intensity than vowels before voiced codas (Archer et al., 2016; House & Fairbanks, 1953). As a correlate of coda voicing, intensity is a potential source of differences in perceived vowel duration based on coda voicing. Given that intensity contour has been observed to influence perceived duration in nonlinguistic stimuli (Grassi & Darwin, 2006; Schlauch et al., 2001), it is likely to similarly influence perceived duration within vowels.

Experiment 3 tests the influence of spectral tilt on perceived vowel duration, as well as the influence of the original coda produced with the vowel and the effect of voiced and voiceless spliced endings. Spectral tilt is higher before non-glottalized voiceless codas than before voiced codas (Coleman, 2003; Sanker, 2019a), and lower before glottalized codas (Chong & Garellek, 2018; Seyfarth & Garellek, 2018). As a correlate of coda voicing, spectral tilt is a potential source of differences in perceived vowel duration based on coda voicing. Given that intensity has been observed to influence perceived duration in nonlinguistic stimuli (Berglund et al., 1969; Goldstone & Lhamon, 1974), and perceived loudness depends on frequency (Fletcher & Munson, 1933; Robinson & Dadson, 1956), spectral tilt is likely to influence perceived vowel duration. While Experiment 2 tested

spliced endings of a single place of articulation, Experiment 3 includes two places of articulation (bilabial and alveolar). Spectral tilt differs by place of articulation for voiceless codas in English, with greater glottalization and thus lower spectral tilt before /t/ than before /p/ (Chong & Garellek, 2018; Seyfarth & Garellek, 2015); different expectations for spectral tilt in each environment might influence how it is perceived.

Experiment 4 tests whether the effect of the original coda is specific to English stimuli or is also present in stimuli produced by a speaker of Telugu, a language without substantial voicing-conditioned vowel duration. Because of voicing-conditioned vowel duration in English, it is unclear whether differences in perceived duration based on the original coda in Experiments 1–3 are due to different effects of lengthening and shortening vowels in the duration manipulation process or are due to effects of the voicing of the original coda. In contrast to English, vowels in Telugu have similar length before voiced and voiceless consonants (Reddy, 1988; Sanker, 2018). Telugu has contrastive vowel length, so using base recordings with phonologically long vowels made it possible to produce nearly all stimuli by shortening the original vowel.

2.1. Stimuli

Stimuli for all experiments were elicited individually in randomized order with PsychoPy (Pierce, 2007) as part of a larger set of items, 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 decision was made to use naturally produced items as the basis of manipulations, rather than trying to force speakers to maintain consistency in duration or other characteristics. Because the items in the original recordings had the variability of naturally produced items, controlling vowel duration in the final stimuli required some variation in the manipulation used for each item.

In each experiment, there were an equal number of stimuli with vowels that had been produced with a voiced coda and stimuli with vowels that had been produced with a voiceless coda. To create the stimuli, the codas, including the closure, release burst, and the formant transition to the coda, were removed from the original recordings. The end of the steady state of the vowel was defined as the 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 vowels were 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 1a**. An example of what the ultimate stimuli looked like is presented in **Figure 1b**. A secondary pilot study demonstrated that the resulting stimuli were perceived as not having codas, except when new codas were spliced on. The onsets were consistent within each experiment, and were left intact unless otherwise specified. The stimuli presented in Experiments 1 and 4 were isolated vowels, whereas in Experiments 2 and 3 the stimuli had a CVC structure.

Vowel duration was manipulated to create a 10-step continuum for each experiment, by copying or removing full glottal cycles in the middle of each vowel. The duration difference between steps was thus constrained by the length of the glottal cycles. Because the base recordings were naturally produced items, F0 was not identical across them. Stimuli were selected and prepared so that the durations of stimuli for each step were nearly identical across items within each experiment, even though this resulted in the continuum range not being identical across experiments: 130 ms to 252 ms in Experiment 1, 120 ms to 212 ms in Experiment 2, 122 ms to 230 ms in Experiment 3, and 130 ms to 252 ms in Experiment 4.

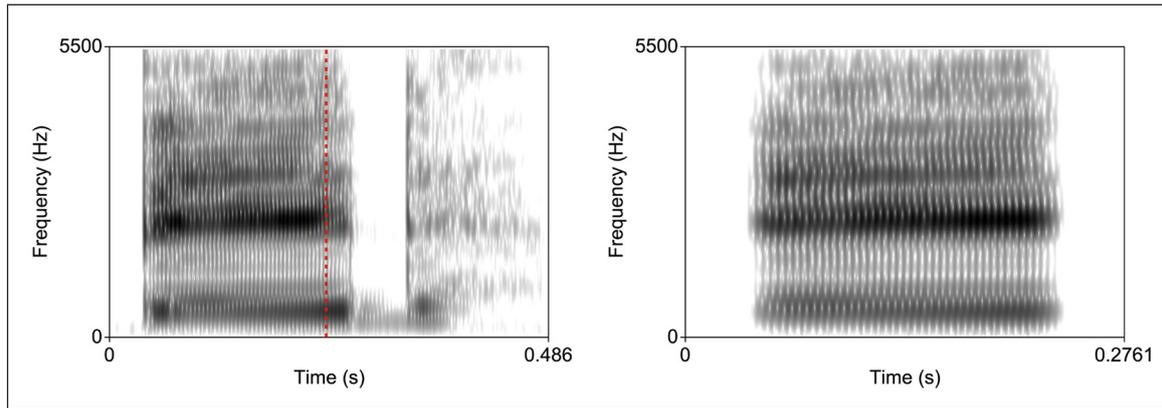


Figure 1: Example spectrograms: (a) Spectrogram of *bid*, before manipulations. The dotted line indicates the calculated end of the steady-state portion of the vowel; this is where the vowel of the stimulus ended. (b) Example of a stimulus made from this recording, used in Experiment 1 (duration step 6, unchanged F1).

Prioritizing naturalness meant that the duration of the original naturally produced vowels was also variable, which resulted in non-identical manipulations for each item. The primary difference was based on coda voicing, though there was also a smaller amount of variation across vowel qualities and across vowels in the same coda voicing environment in different experiments. Because of the voicing-conditioned vowel duration differences in English, vowels produced before voiced codas had a natural duration around duration step 9 and were shortened to produce most voicing steps, while vowels produced before voiceless codas had a natural duration around duration step 4 and were lengthened to produce most voicing steps. With the Telugu stimuli in Experiment 4, all vowels in the nonce words were phonologically long and nearly all of them were manipulated only by removing glottal cycles; the one exception was /i/ before voiceless /t/, for which the longest duration step had to be created by copying existing glottal cycles.

2.1.1. Stimuli for Experiment 1

Stimuli for Experiment 1 were based on recordings of a female American English speaker producing six words with the onset /b/, followed by each of 3 vowel qualities (/æ, ε, ɪ/) and 2 coda environments: voiced /d/ and voiceless /t/. In order to avoid the possibility that the duration of the transition from the onset might be used as a cue for duration, which could produce differences across vowel qualities, the onset was removed following the same procedures described above for codas, and each vowel was given a 30 ms fade-in.¹

Within each vowel quality, F1 was manipulated to create three items: the vowel with the naturally produced F1, a vowel with the F1 increased 60 Hz + 6% above the naturally produced F1, and a vowel with the F1 decreased 60 Hz + 6% below the naturally produced F1. The manipulation was chosen based on previous evidence that the just noticeable difference in formants depends both on the absolute size of the difference and the relative size of the difference, as well as varying substantially across individuals (Kewley-Port & Watson, 1994; Mermelstein, 1978). The selected level was reliably larger than the typical JND for each vowel while not being so large that the item was shifted into the prototypical region of a different American English vowel; if the manipulation were too small, it might not produce any effect, while too large a manipulation would obscure possible differences

¹ The original onsets in Experiments 2 and 3 were not removed; unlike Experiment 1, these experiments did not involve comparisons across vowel qualities, so possible effects of the transition to different vowel qualities were not a concern.

between effects of vowel category and effects of gradient F1. There were also 10 vowel duration steps, as described above. This produced a total of 180 items, each of which was heard a single time by each listener.

The following tables present summary acoustic measurements of the stimuli, separated by the original coda environment. As discussed above, the voicing of the coda in production influences vowels in a range of ways, and some of these might contribute to perceived duration. Although the data is separated by the voicing of the original coda, the measurements are from the stimuli used in the experiment itself, in which these codas had been removed. F0 change and intensity change were measured as the difference between the mean value in the final quarter of the vowel and the first quarter of the vowel. F1 was z-scored in experiments which used multiple vowel qualities, to allow pooled measurements across vowel qualities. Spectral tilt was measured in the final quarter of the vowel, given that the effect of coda voicing is often local. All of the measurements were averages within the specified interval, either the whole vowel, the first 25% of the vowel, or the final 25% of the vowel.

Table 1 presents a summary of the acoustic characteristics of the stimuli used in Experiment 1, which may contribute to differences in perceived duration between vowels originally produced before voiced codas and voiceless codas. A further breakdown of acoustic information by stimulus item is provided in the Appendix.

2.1.2. Stimuli for Experiment 2

Stimuli for Experiment 2 were based on recordings of a female American English speaker producing the words *badge* and *batch*, i.e., a consistent /bæ/ onset and vowel followed by two coda environments differing only in voicing.

Each vowel was manipulated to have three intensity contours: rising (10.8 dB rise), falling (11.1 dB fall), and approximately level (1.1 dB rise). Each vowel was spliced with two different endings: voiceless /k/ and voiced /g/. There were also 10 vowel duration steps, as described above. This produced a total of 120 items, each of which was heard a single time by each listener.

Table 2 presents a summary of the acoustic characteristics of the stimuli used in Experiment 2. The data is separated by the voicing of the original coda, because this

Table 1: Acoustic characteristics of vowels in the Experiment 1 stimuli, by the original coda's voicing.

	f0 mean	f0 change (Q4-Q1)	intensity change (Q4-Q1)	F1 (z-scored)	spectral tilt (Q4)	jitter	HNR
Voiced	213 Hz	4.6 Hz	1.2 dB	-0.35	-12.0 dB	0.6%	14.3
Voiceless	232 Hz	8.5 Hz	-0.6 dB	0.35	-9.2 dB	0.6%	14.7

Table 2: Acoustic characteristics of vowels in the Experiment 2 stimuli, by the original coda's voicing.

	f0 mean	f0 change (Q4-Q1)	intensity change (Q4-Q1)	F1	spectral tilt (Q4)	jitter	HNR
Voiced	194 Hz	-7.6 Hz	2.0 dB	883 Hz	2.9 dB	1.3%	8.7
Voiceless	198 Hz	-7.9 Hz	-1.5 dB	1057 Hz	6.1 dB	1.5%	8.1

original production environment has an impact on perceived duration; these acoustic measurements provide possible sources of those perceptual differences. However, the measurements are from the stimuli used in the experiment itself, in which these codas had been removed and new codas had been spliced on.

2.1.3. Stimuli for Experiment 3

Stimuli for Experiment 3 were based on recordings of a female American English speaker producing the words *cub* and *cup*, i.e., a consistent /kʌ/ onset and vowel followed by two coda environments that differed only in voicing.

Each vowel was manipulated to have three spectral tilts, using the Praat Filter (formula) function to recalculate intensities relative to frequency: high (10.5 dB), moderate (7.0 dB), and low (3.2 dB). The filter formula used to create high spectral tilt items was: $self / (1 + x/100)^{0.8}$, and the filter formula used to create low spectral tilt items was: $self * (1 + x/100)^{-0.8}$. The moderate spectral tilt items preserved the natural spectral tilt of the vowel. The intensity of the resulting items was then normalized so that they all had the same mean intensity. Note that the manipulation was relative to the original spectral tilt, so it preserved voicing-conditioned spectral tilt differences.

Each vowel was spliced with four different endings: voiceless /p/ and /t/ and voiced /b/ and /d/. There were also 10 vowel duration steps, as described above. This produced a total of 240 items, each of which was heard a single time by each listener.

Table 3 presents a summary of the acoustic characteristics of the vowels in the stimuli used in Experiment 3. The data is separated by the voicing of the original coda, because this original production environment has an impact on perceived duration; these acoustic measurements provide possible sources of those perceptual differences. However, the measurements are from the stimuli used in the experiment itself, in which these codas had been removed and new codas had been spliced on.

2.1.4. Stimuli for Experiment 4

Stimuli for Experiment 4 were based on recordings of a female Telugu speaker producing VC nonce words. The stimulus vowels came from 4 coda environments: voiced /d/ and /g/ and voiceless /t/ and /k/, preceded by 3 vowel qualities: /a, i, u/. In order to reduce variability across stimuli, vowels were defined as the onset of voicing and given a 30 ms fade-in; before this manipulation, some vowels were preceded by voiceless noise, which could have complicated the evaluation of vowel duration. There were also 10 vowel duration steps, as described above. This produced a total of 120 items, each of which was heard a single time by each listener.

Table 4 presents a summary of the acoustic characteristics of the vowels in the stimuli used in Experiment 4. The data is separated by the voicing of the original coda, because this original production environment has an impact on perceived duration; these acoustic measurements provide possible sources of those perceptual differences. However, the

Table 3: Acoustic characteristics of vowels in the Experiment 3 stimuli, by the original coda's voicing.

	f0 mean	f0 change (Q4-Q1)	intensity change (Q4-Q1)	F1	spectral tilt (Q4)	jitter	HNR
Voiced	168 Hz	-19.6 Hz	0.7 dB	696 Hz	2.8 dB	1.0%	9.0
Voiceless	167 Hz	-24.6 Hz	-0.4 dB	775 Hz	5.0 dB	1.4%	6.6

measurements are from the stimuli used in the experiment itself, in which these codas had been removed.

As mentioned above, one of the key characteristics of Telugu that is relevant for this experiment is that vowel durations do not differ substantially based on the voicing of following consonants (Reddy, 1988; Sanker, 2018). There is similarly little difference in the original durations of the vowels that were used to make stimuli for Experiment 4. **Table 5** presents a summary of the naturally produced durations of vowels, by vowel quality and coda voicing. Each value is the average of the durations for two items, one with a dental coda and one with a velar coda.

2.2. Experiment design

In each experiment, participants were instructed to make decisions about vowel duration. The instructions were given verbally: “You are going to hear vowels 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 [æ] or a long vowel like [æ:].” They were then asked whether the task was clear, and were invited to ask questions if they wanted clarification; all of them confirmed their understanding. They did not receive any feedback to train them on a particular duration divide.

The experiment was run in PsychoPy. Listeners identified each vowel as ‘short’ or ‘long,’ indicated with the arrow keys on a computer keyboard. The arrow keys associated with ‘long’ and ‘short’ were balanced across participants. The experiment was self timed; a new trial began only after a response was given.

The order of presentation of items was randomized. When vowel quality varied across items (Experiments 1 and 4), items were blocked by vowel quality and randomization was within blocks, in order to facilitate within-category evaluations of duration. When vowel quality was consistent across items (Experiments 2 and 3), randomization was across all items.

Statistical results are from mixed effects models, calculated with the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015). *P*-values were calculated by the lmerTest package (Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2015). Responses with latencies shorter than 250 ms or longer than 5 s were excluded from analysis (between 1.0% and 1.4% of the data in each experiment).

Table 4: Acoustic characteristics of vowels in the Experiment 4 stimuli, by the original coda’s voicing.

	f0 mean	f0 change (Q4-Q1)	intensity change (Q4-Q1)	F1 (z-scored)	spectral tilt (Q4)	jitter	HNR
Voiced	229 Hz	42.4 Hz	4.3 dB	-0.22	3.9 dB	1.1%	11.3
Voiceless	230 Hz	35.3 Hz	3.1 dB	0.22	4.3 dB	0.9%	11.2

Table 5: Original durations of vowels used in Experiment 4, by vowel quality and the original coda’s voicing.

	a	i	u
Voiced	292 ms	229 ms	216 ms
Voiceless	287 ms	227 ms	223 ms

There were 96 total participants, all native speakers of American English: 24 in Experiment 1 (6 male, 18 female; mean age 21.3), 24 in Experiment 2 (4 male, 20 female; mean age 21.5), 24 in Experiment 3 (3 male, 21 female; mean age 19.8), and 24 in Experiment 4 (7 male, 17 female; mean age 22.0).

2.3. Hypotheses and predictions

Hypothesis 1a: Experiment 1 tests the effect of vowel category on perceived vowel duration. Listeners' awareness of the relationship between vowel height and duration in production is likely to result in compensation for expected duration across categories. As has been found in previous work, listeners are likely to perceive high vowels as longer than low vowels (Gussenhoven, 2007; Wang et al., 1976).

Hypothesis 1b: Experiment 1 also tests the effect of F1 within vowel categories on perceived vowel duration. If the relationship between vowel height and duration is a continuum, the same effect observed across categories should be apparent within categories: Vowels with a lower F1 would be perceived as longer than vowels with higher F1. Such a relationship could be responsible for greater perceived duration of vowels produced before voiced codas, given that coda voicing results in lower F1 in preceding vowels (Lisker, 1986; Summers, 1987). A lack of effect would suggest that listeners learn the relationship between F1 and duration by category, and that it is not an automatic perceptual effect.

Hypothesis 2: Experiment 2 tests the effect of intensity contour on perceived vowel duration. As has been observed for non-linguistic stimuli (Grassi & Darwin, 2006; Schlauch et al., 2001), rising intensity is likely to increase listeners' perceived duration of a segment, and falling intensity is likely to decrease perceived duration, resulting in more identifications of vowels as long when they have a rising intensity contour, and fewer identifications of vowels as long when they have a falling intensity contour. Such a relationship could be responsible for greater perceived duration of vowels produced before voiced codas, given that coda voicing results in higher final intensity in preceding vowels (Archer et al., 2016; House & Fairbanks, 1953).

Hypothesis 3: Experiment 3 tests the effect of spectral tilt on perceived vowel duration. Spectral tilt may increase listeners' perceived duration of a segment, given that perceived intensity increases with frequency (Fletcher & Munson, 1933; Robinson & Dadson, 1956), and greater intensity increases perceived duration (Berglund et al., 1969; Goldstone & Lhamon, 1974). Higher spectral tilt is likely to decrease perceived intensity and subsequently perceived duration, resulting in a negative relationship between spectral tilt and the proportion of 'long' responses. Such a relationship could be responsible for greater perceived duration of vowels produced before voiced codas, given that coda voicing results in lower spectral tilt than is found before non-glottalized voiceless codas (Coleman, 2003; Sanker, 2019a).

Hypothesis 4: All four experiments test the effects of the original coda produced with a vowel on the perceived duration of that vowel. Some of the acoustic differences between vowels produced in the two coda voicing environments could influence perceived duration; listeners may perceive vowels as longer if they were produced with voiced codas (Sanker, 2019b). However, given the substantial voicing-conditioned vowel duration effects within English, the duration manipulations were different for vowels from each coda voicing environment in Experiments 1–3: English vowels produced before voiceless codas are shorter than vowels produced before voiced codas, so the duration continuum required lengthening for most vowels from the voiceless environment and shortening for most vowels from the voiced environment. Experiment 4 tests whether there is an effect of the original coda voicing environment when stimuli come from Telugu, a language without

significant voicing-conditioned vowel duration; an effect of the original coda environment in this experiment would suggest that the results in the other experiments are not merely the result of the different duration manipulations for vowels from each environment, but are a result of how coda voicing influences characteristics of preceding vowels. If there is an effect of the original coda environment in Experiments 1–3 but not in Experiment 4, this would suggest that the different duration manipulations necessary for vowels from each environment influenced perceived duration.

Hypothesis 5: Experiments 2 and 3 test the effect of a spliced-in coda on perceived vowel duration. The voicing of the coda presented with each vowel could influence listeners' perception of duration: Because vowels are usually longer before voiced codas, listeners may compensate for the expected duration in the environment, resulting in longer perceived vowel duration when vowels are presented with voiceless codas.

3. Results

3.1. Experiment 1: F1 and vowel quality

Experiment 1 tests the influences of F1 and vowel quality on perceived vowel duration. **Table 6** presents the summary of a mixed effects logistic regression model for the 'long' responses to each item in Experiment 1. The fixed effects were vowel duration step; vowel quality (/æ, ε, ɪ/); F1 step within the vowel; and voicing of the original coda (voiced, voiceless). There was a random intercept for participant, and there were no random slopes.²

Duration was a significant predictor of responses: Longer vowels were more often identified as long. The relationship between actual duration and perceived duration is illustrated in **Figures 2** and **3**.

Vowel quality was a significant factor in responses. /ɪ/ elicited more 'long' responses than /ε/, though /ε/ and /æ/ did not differ significantly. **Figure 2** illustrates responses separated by vowel quality. The effect of vowel quality is weaker at long durations, likely because the longest duration steps approach the ceiling of perceived vowel duration; if a listener perceives all vowels as long at that duration, there is no room for an effect of

Table 6: Regression model for 'long' responses, Experiment 1. *Reference Levels: Vowel = /ε/; OrigCoda = Voiced.*

	β	SE	z value	p value
(Intercept)	-2.95	0.213	-13.8	<0.001***
Duration Step	0.505	0.0162	31.3	<0.001***
Vowel /ɪ/	0.359	0.0936	3.84	<0.001***
Vowel /æ/	-0.061	0.0934	-0.652	0.514
F1	0.0134	0.0467	0.288	0.774
OrigCoda Voiceless	-0.17	0.0763	-2.23	0.026*

² Maximizing random effects structure, as suggested by Barr, Levy, Scheepers, and Tily (2013) has some drawbacks, which are laid out by Matuschek, Kliegl, Vasishth, Baayen, and Bates (2017): Extensive random effects reduce the chance of Type I errors, but substantially increase Type II errors, as well as frequently producing models that fail to converge because of overparameterization. A parsimonious random effects structure was chosen based on the lack of clear theoretical reason to expect most of the fixed effects to exhibit variation across individuals. Individual differences in sensitivity to vowel duration might justify a random slope for duration step. However, this slope was excluded in order to maintain parallel random effects structure across models; it caused some models to fail to converge and its inclusion never altered the results in the models which did converge.

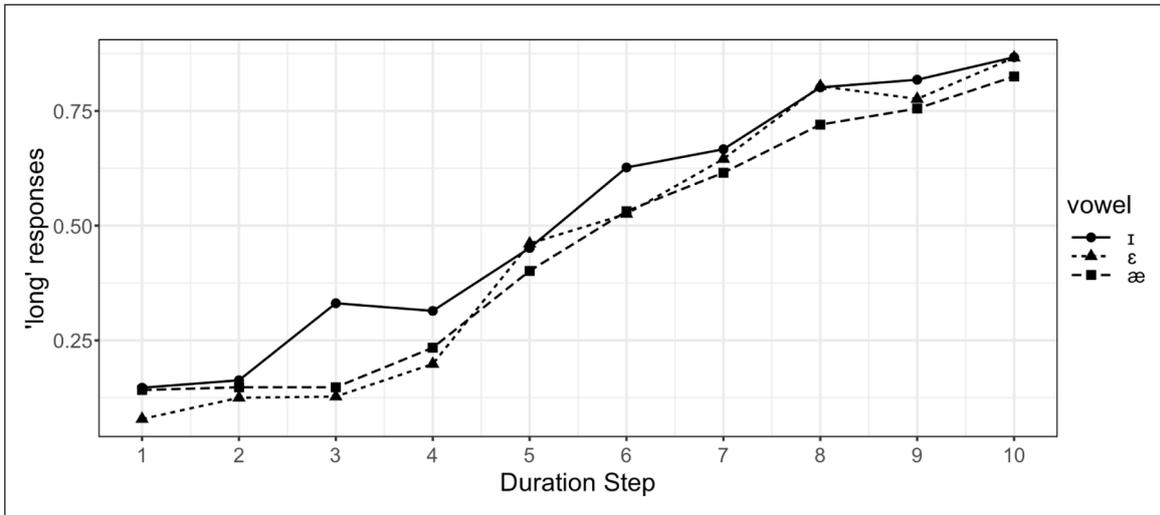


Figure 2: Proportion of ‘long’ responses in Experiment 1, by duration step and vowel. Based on the raw data, not the output of the regression model, and pooled across participants.

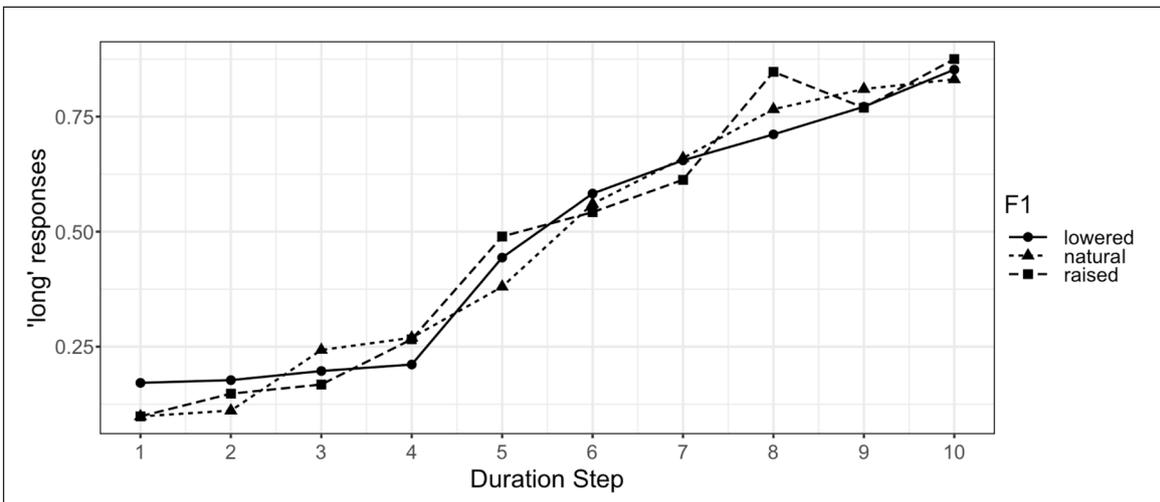


Figure 3: Proportion of ‘long’ responses in Experiment 1, by duration step and within-category F1. Based on the raw data, not the output of the regression model, and pooled across participants.

vowel quality. Including the interaction between vowel quality and duration step does improve the model ($\chi^2 = 8.6, df = 2, p = 0.0136$); the model with this interaction can be found in the Appendix, but the simpler model is kept here, as it is not clear that the interaction is best modelled as linear and the interaction does not change the main effects.

Within vowel category, F1 was not a significant predictor of ‘long’ responses. **Figure 3** illustrates responses separated by F1 manipulation. In the regression model, F1 is numbered from 1 (lowest F1) to 3 (highest F1), rather than using absolute values, to allow pooling across the vowel qualities and original coda environments.

The voicing of the coda which had originally been produced with the vowel was also a significant factor. Vowels from the environment of a voiceless coda were identified as long less frequently than vowels from the environment of a voiced coda.

Additional figures are provided in the Appendix, to help visualize the factors that are not included in the figures here.

3.2. Follow-up to Experiment 1: Perception of vowel quality

A brief follow-up study was conducted to test the perceived category of each vowel and check whether the F1 manipulations or the duration manipulations influenced perceived vowel quality. Effects on perceived vowel quality could influence perceived duration, and might explain the lack of clear effect of sub-categorical F1 manipulations on perceived duration.

Twenty-five native American English speakers (1 male, 24 female; mean age 21.9) completed a forced choice task about vowel quality. Participants heard a subset of the stimuli used in Experiment 1: Each of the three vowel categories (/æ, ε, ɪ/) from the two coda environments (t, d), with each of the three manipulations (natural F1, raised F1, lowered F1), at three durations (130 ms, 197 ms, 252 ms), for a total of 54 items. These items were mixed with 80 filler trials of other vowel qualities, and were presented in randomized order. For each trial, listeners indicated the vowel quality by identifying the matching vowel in an array of monosyllabic English words with each vowel in the same environment (*beat, bit, bet, bat*). Each response was associated with an arrow key and presented in the corresponding position on the screen. The experiment was run in PsychoPy and was self-timed.

The results are based on Chi-Square tests of independence, using the counts of each response based on the three F1 manipulations and the three duration manipulations.

Figures 4–5 present a summary of responses for each vowel, by F1 manipulation and duration manipulation. Recall that these response labels are not the ones which listeners used; they identified vowel quality by matching it with English words.

While stimuli are described in reference to the originally produced vowel they came from, it is important to note that the vowels had been manipulated in several ways that

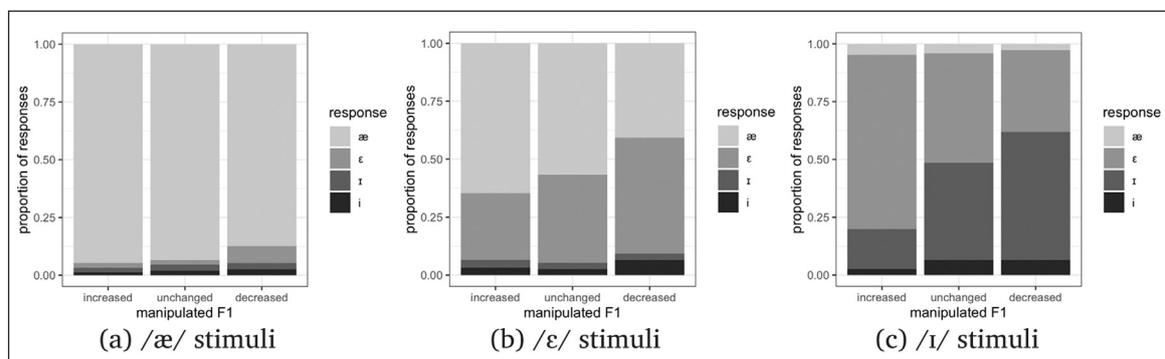


Figure 4: Effects of F1 manipulations on identifications of each stimulus vowel.

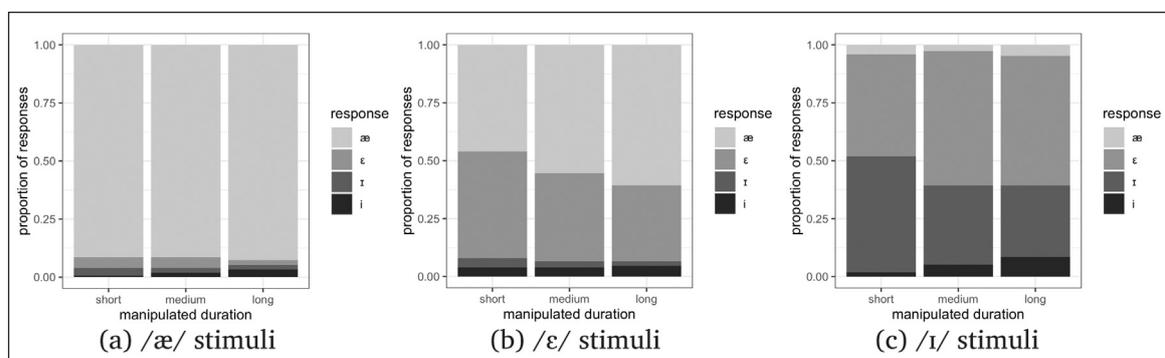


Figure 5: Effects of duration manipulations on identifications of each stimulus vowel.

might be expected to alter their perceived quality; this follow-up study aims to evaluate how these potentially ambiguous vowels were being perceived. F1 manipulations are extremely likely to influence perceived vowel quality, even though they were not intended to substantially shift the vowel category. Changes in vowel duration and the elimination of the coda could also alter the perceived quality of the vowels. Thus, the responses should not be interpreted as measuring accuracy, but rather testing how ambiguous the identity of these vowels was and illustrating the contribution of F1 and vowel duration to perceived vowel quality.

The stimuli made from /æ/ were consistently identified as /æ/, regardless of the F1 manipulation. However, identifications of /ε/ stimuli were influenced by the F1 manipulation condition, and responses were mixed between /æ/ and /ε/. /ε/ responses for /ε/ stimuli increased with lower F1: $\chi^2(6, N = 450) = 20.5, p = 0.0022$.

Identifications of /ɪ/ stimuli were also influenced by the F1 manipulation condition; responses were mixed between /ε/ and /ɪ/. /ɪ/ responses for /ɪ/ stimuli increased with lower F1: $\chi^2(6, N = 450) = 57.0, p < 0.001$.

Variable perception of the quality of the /ε/ and /ɪ/ tokens could have obscured effects of within-category F1 variation on perceived vowel duration in Experiment 1. However, vowels in Experiment 1 were blocked by vowel quality, which probably reduced listeners' uncertainty about identity, while vowel qualities in this vowel identification task were mixed with no blocking.

The perception of vowel height was also influenced by vowel duration. Identifications of /æ/ stimuli were not influenced by duration, but duration did influence identifications of the other vowels. /ɪ/ stimuli were more likely to be identified as mid at longer durations, and also exhibited a small increase in /i/ responses: $\chi^2(6, N = 450) = 18.7, p = 0.0047$. /ε/ stimuli exhibited a similar trend towards more /æ/ identifications at longer durations, though the effect was not significant: $\chi^2(6, N = 450) = 7.7, p = 0.26$.

The effect of vowel duration on perceived vowel height seems to depend on ambiguity of the formants. As the responses to /æ/ stimuli indicate, the duration has very little effect on decisions when the formants are unambiguously within one vowel category. For /ε/ and /ɪ/, the vowel quality was ambiguous in all of the F1 manipulation conditions, as seen in **Figure 4**. However, the effect of vowel duration is distinct from the effect of F1; there were an equal number of tokens with each F1 manipulation at each vowel duration, so F1 is consistent across durations. The Appendix presents figures isolating responses for each combination of duration and F1 manipulation.

The effect of duration on perceived vowel height has also recently been demonstrated by Kim and Clayards (2019), with English /ε/ and /æ/. More work has looked for an effect of duration on English speakers' perception of vowel tenseness, particularly with /i/ versus /ɪ/, given the correlation in production. English speakers largely ignore duration as a cue for tenseness (Casillas, 2015), though an effect of vowel duration on perceived tenseness can be found when listeners have a heavy cognitive load (Gordon, Eberhardt, & Rueckl, 1993).

If the lack of effect on within-category F1 in Experiment 1 is because the manipulations influenced perceived vowel quality, then an effect might be apparent when restricting the data to /æ/ items, which were consistently perceived as /æ/. Thus, the same analysis presented above was replicated just for /æ/ tokens. **Table 7** presents the summary of a mixed effects logistic regression model for the 'long' responses to /æ/ items in Experiment 1. The fixed effects were vowel duration step; F1 step within the vowel; and voicing of the original coda (voiced, voiceless). There was a random intercept for participant, and there were no random slopes.

Even when restricting the data to items for which perceived vowel quality was not influenced by the F1 manipulation, there was no effect of F1 on perceived vowel duration. Thus, the results still provide no evidence for any effect of within-category F1 on perceived vowel duration.

3.3. Experiment 2: Intensity

Experiment 2 tests the effects of intensity contour on perceived vowel duration, as well as the effect of voiced and voiceless spliced endings. **Table 8** presents the summary of a mixed effects logistic regression model for the ‘long’ responses to each item in Experiment 2. The fixed effects were vowel duration step; intensity contour (rising, falling, level); voicing of the original coda (voiced, voiceless); and the voicing of the spliced ending (voiced, voiceless). As in Experiment 1, there was a random intercept for participant, and there were no random slopes.

Duration was a significant predictor of responses, as in Experiment 1: Longer vowels were more likely to be identified as long. The relationship between actual duration and perceived duration is illustrated in **Figure 6**.

Intensity contour was a significant factor in responses. Rising intensity items elicited significantly more ‘long’ responses than level intensity items, and falling intensity items elicited significantly fewer. The effects of intensity contour on perceived duration are illustrated in **Figure 6**.

The voicing of the coda which had originally been produced with the vowel was a significant factor. Vowels originally produced before voiced codas elicited more ‘long’ responses.

The voicing of the spliced ending was also a significant predictor; listeners were significantly more likely to identify a vowel as long when it was presented with a voiceless coda, suggesting an awareness of the phonological relationship between vowel duration and coda voicing, and compensation for the contextually expected vowel duration.

Table 7: Regression model for ‘long’ responses to /æ/ items, Experiment 1. *Reference Levels: OrigCoda = Voiced.*

	β	SE	z value	p value
(Intercept)	-2.95	0.342	-8.63	<0.001***
Duration Step	0.525	0.0299	17.6	<0.001***
F1	-0.0347	0.0838	-0.413	0.679
OrigCoda Voiceless	-0.334	0.137	-2.44	0.0148*

Table 8: Regression model for ‘long’ responses, Experiment 2. *Reference Levels: Intensity = Level; OrigCoda = Voiced; Ending = Voiced.*

	β	SE	z value	p value
(Intercept)	-2.14	0.155	-13.8	<0.001***
Duration Step	0.314	0.0164	19.2	<0.001***
Intensity Rising	0.397	0.105	3.78	<0.001***
Intensity Falling	-0.532	0.106	-5.02	<0.001***
OrigCoda Voiceless	-0.47	0.0865	-5.43	<0.001***
Ending Voiceless	1.24	0.0887	14.0	<0.001***

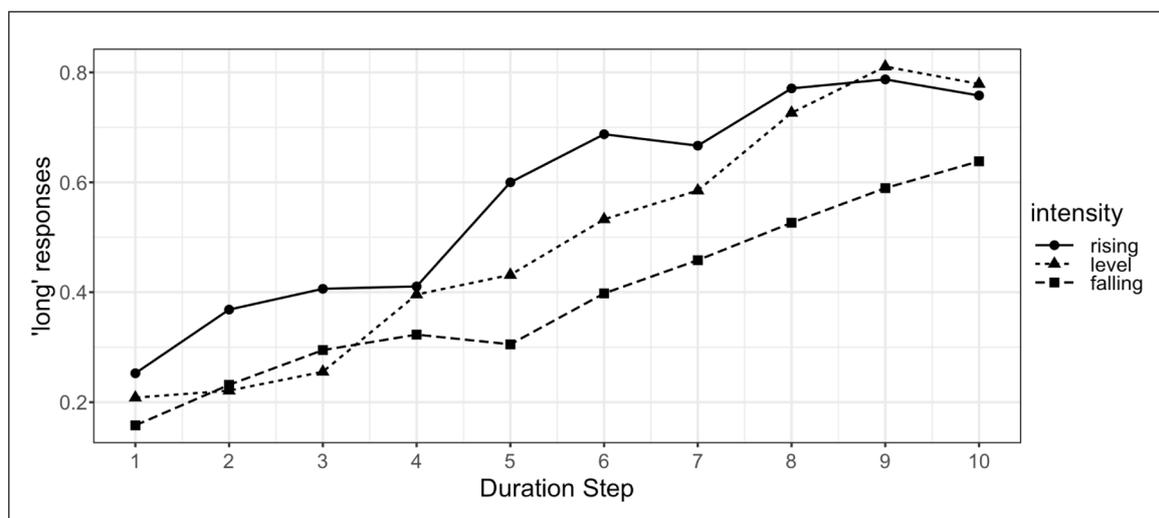


Figure 6: Proportion of 'long' responses in Experiment 2, by duration step and intensity contour. Based on the raw data, not the output of the regression model, and pooled across participants.

An additional model was run which included an interaction between intensity and spliced ending, to check whether these two manipulations might impact each other. This interaction did not improve the model ($\chi^2 = 2.1$, $df = 2$, $p = 0.351$), and similarly is not a significant factor in the outputs of such models, so it was not included.

3.4. Experiment 3: Spectral tilt

Experiment 3 tests the effects of spectral tilt on perceived vowel duration, as well as the effect of voiced and voiceless spliced endings. **Table 9** presents the summary of a mixed effects logistic regression model for the 'long' responses to each item in Experiment 3. The fixed effects were vowel duration step; spectral tilt (high, low, moderate); voicing of the original coda (voiced, voiceless); and voicing of the spliced ending (voiced, voiceless). As in Experiment 1, there was a random intercept for participant, and there were no random slopes.

Duration was a significant predictor of responses, similar to the results in the previous experiments: Longer vowels were more likely to be identified as long. The relationship between actual duration and perceived duration is illustrated in **Figure 7**.

Spectral tilt was a significant predictor of responses. Relative to vowels with moderate spectral tilt, vowels with high spectral tilt were significantly less likely to be perceived as long. The effects of spectral tilt are illustrated in **Figure 7**.

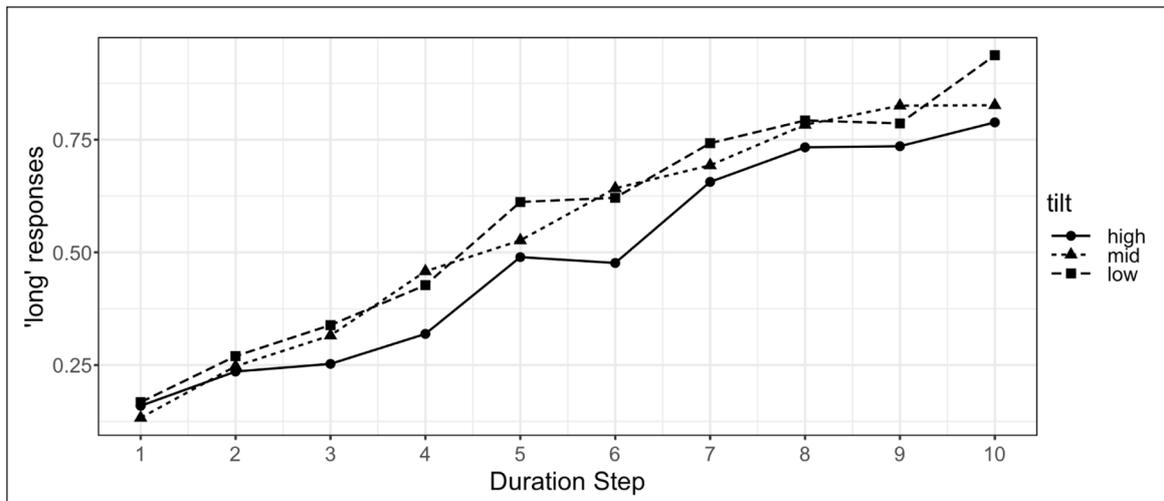
The voicing of the coda which had originally been produced with the vowel was a significant factor. Vowels originally produced before voiced codas elicited more 'long' responses.

The voicing of the spliced ending was also a significant predictor; listeners were significantly more likely to identify a vowel as long when it was presented with a voiceless coda. Although the stimuli included four endings (/b, d, p, t/), separating this factor into the four specific endings did not produce a model with a better fit than one which instead grouped the endings based on voicing ($\chi^2 = 0.431$, $df = 2$, $p = 0.806$), so the simpler model is the one presented here. The model with separate endings can be found in the Appendix.

An additional model was run which included an interaction between spectral tilt and spliced ending, to check whether these two manipulations might impact each other. This interaction did not improve the model ($\chi^2 = 0.104$, $df = 2$, $p = 0.949$), and similarly is

Table 9: Regression model for ‘long’ responses, Experiment 3. *Reference Levels: Spectral tilt = Moderate; OrigCoda = Voiced; Ending = Voiced.*

	β	SE	z value	p value
(Intercept)	-2.0	0.16	-12.4	<0.001***
Duration Step	0.409	0.0125	32.6	<0.001***
Spectral Tilt High	-0.333	0.0766	-4.35	<0.001***
Spectral Tilt Low	0.132	0.0768	1.72	0.0856
OrigCoda Voiceless	-0.21	0.0626	-3.36	<0.001***
Ending Voiceless	0.2	0.0626	3.19	0.00143**

**Figure 7:** Proportion of ‘long’ responses in Experiment 3, by duration step and spectral tilt. Based on the raw data, not the output of the regression model, and pooled across participants.

not a significant factor in the outputs of such models, so it was not included. An interaction between spectral tilt and the specific ending also did not improve the model ($\chi^2 = 4.66$, $df = 8$, $p = 0.793$).

3.5. Experiment 4: Coda voicing

Experiment 4 tests whether the effect of the original coda is specific to English stimuli or is also present in stimuli produced by a speaker of Telugu, a language without significant voicing-conditioned vowel duration. **Table 10** presents the summary of a mixed effects logistic regression model for the ‘long’ responses to each item in Experiment 4. The fixed effects were vowel duration step; vowel quality (/a, i, u/); and voicing of the original coda (voiced, voiceless). As in Experiment 1, there was a random intercept for participant, and there were no random slopes.

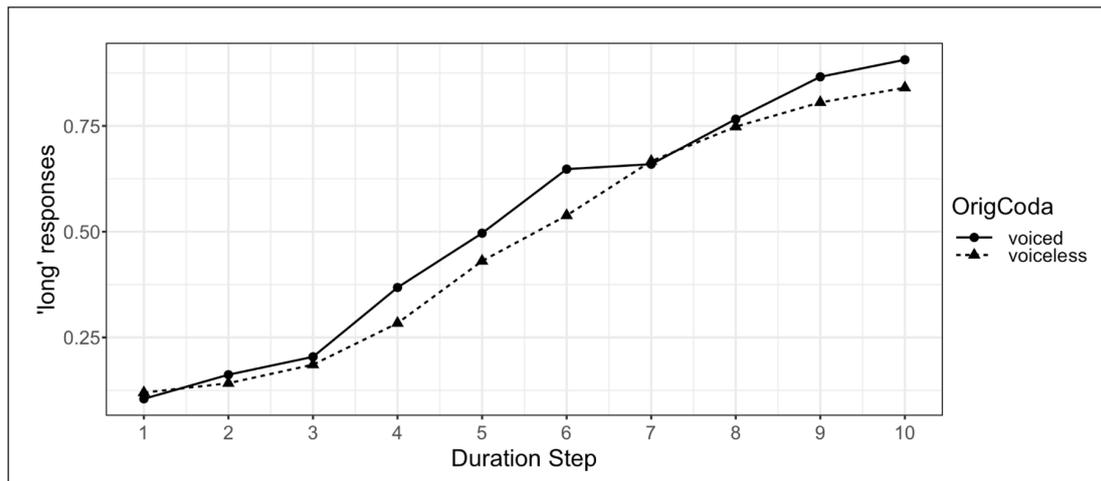
Duration was a significant predictor of responses, similar to the results in the previous experiments. The relationship between actual duration and perceived duration is illustrated in **Figure 8**.

Vowel quality was not a significant predictor of responses.

The voicing of the coda which had originally been produced with the vowel was a significant factor. Vowels originally produced before voiced codas elicited more ‘long’ responses. The effects of coda voicing are illustrated in **Figure 8**.

Table 10: Regression model for ‘long’ responses, Experiment 4. *Reference Levels: Vowel = a; OrigCoda = Voiced.*

	β	SE	z value	p value
(Intercept)	-2.71	0.173	-15.7	<0.001***
Duration Step	0.495	0.0192	25.8	<0.001***
Vowel /i/	0.119	0.112	1.06	0.288
Vowel /u/	0.175	0.112	1.56	0.118
OrigCoda Voiceless	-0.246	0.0917	-2.68	0.00735**

**Figure 8:** Proportion of ‘long’ responses in Experiment 4, by duration step and original coda voicing. Based on the raw data, not the output of the regression model, and pooled across participants.

4. Discussion

The results of these experiments demonstrate several factors that influence perceived vowel duration: Vowel height, intensity contour, spectral tilt, and coda voicing. Effects of vowel category and the voicing of the coda spliced onto the vowel indicate characteristics which have duration as part of the representation; listeners expect certain durations, and compensate for those expectations. The effects of spectral tilt and intensity contour seem to reflect perceptual influences not dictated by phonological expectations. Moreover, because they are correlates of coda voicing, they provide a possible perceptual pathway for voicing effects on vowel duration: Coda voicing creates these vowel characteristics, they influence perceived vowel duration, and the different perceived vowel durations in each environment could then enter the representation as distinct durations conditioned by coda voicing.

Even though vowel length is not contrastive per se in American English, listeners are sensitive to vowel duration and can make accurate decisions about it, as demonstrated by the strong relationship between actual duration and the number of ‘long’ responses to stimuli in all four experiments. Unlike perception along dimensions that characterize phonological contrasts, duration decisions were close to linear, rather than exhibiting a sharp category boundary.

In Experiment 1, vowel quality influenced perceived duration, consistent with previous work demonstrating that listeners perceive higher vowels as longer (Gussenhoven, 2007; Wang et al., 1976). This study provided a three-way height comparison of vowels

matching in other features (front, unrounded, lax); /ɪ/ was perceived as longer than /ɛ/ or /æ/, which is consistent with listeners being aware of the differences in production and compensating in perception. That is, they expect low vowels to be longer, so a low vowel is shorter relative to its expected duration than a higher vowel of the same duration. However, there was no significant difference between the perceived duration of /æ/ and /ɛ/, which perhaps is due to the greater acoustic overlap between these categories than between /ɪ/ and /ɛ/ in American English (Peterson & Barney, 1952), which is illustrated perceptually in the vowel identification follow-up to Experiment 1. When listeners perceived /ɪ/ as /ɛ/, it would exhibit the same effects of phonological expectations as /ɛ/.

Even though vowel height across vowel categories was a significant predictor of perceived duration, F1 within vowel categories was not a significant predictor of perceived vowel duration. This lack of perceptual correlation within categories is consistent with previous work suggesting that there is no within-category correlation in production between F1 and vowel duration (Toivonen et al., 2015). The effects of vowel height on perceived duration thus seem to be based on learned associations with each vowel category, rather than an effect of absolute F1. Solé and Ohala (2010) provide evidence from production which similarly suggests that listeners encode a categorical relationship between F1 and duration, because the ratio of vowel duration between vowel qualities is consistent at different speech rates.

In Experiment 2, the intensity contour had a large effect on perceived vowel duration. Consistent with results for non-linguistic stimuli (e.g., Grassi & Darwin, 2006; Schlauch et al., 2001), vowels were perceived as longer when they increased in intensity and shorter when they decreased in intensity. With linguistic stimuli, previous work has demonstrated that partial vowel devoicing, which also decreases intensity, leads to shorter perceived vowel duration (Myers & Hansen, 2007). There are larger decreases in intensity before voiceless codas than before voiced codas (Archer et al., 2016; House & Fairbanks, 1953), so the effect of intensity on perceived vowel duration could contribute to voicing-conditioned vowel duration differences.

In Experiment 3, spectral tilt had a significant effect on perceived duration; vowels with higher spectral tilt were perceived as shorter than vowels with lower spectral tilt. This effect suggests a relationship with loudness perception. Higher frequencies are perceived as louder than the same intensity at lower frequencies (Fletcher & Munson, 1933; Robinson & Dadson, 1956), so lower spectral tilt, in which higher frequencies are relatively more intense, could increase perceived loudness. Greater intensity increases perceived duration (Berglund et al., 1969; Goldstone & Lhamon, 1974), so vowels with greater perceived loudness are likely to be perceived as longer. The relationship between spectral tilt and voicing could thus contribute to voicing-conditioned differences in vowel duration.

In Experiments 2 and 3, each vowel was spliced with voiced and voiceless codas. In both experiments, the voicing of the given ending was a significant predictor of responses, with more 'long' responses for vowels presented with voiceless codas. The effect of the coda presented with the vowel is crucially different than the effect of the coda that was originally produced with the vowel, which will be discussed subsequently. Consistent with Sanker (2019b), the effect of the coda presented with the vowel suggests compensation for expected duration in context. Voicing-conditioned vowel duration seems to be part of the phonological representation in English, based on the consistent duration ratio preserved in different stress contexts (de Jong, 2004) and its preservation even when voicing is not produced (Sharf, 1964). Thus, a vowel before a voiceless coda is more likely to be perceived as long relative to the range of expected durations. Listeners' compensation for expected duration provides additional evidence that the difference is phonologized.

Although Experiment 3 included both labial and alveolar spliced endings, there was no evidence that place of articulation influenced perception of vowel duration, either broadly or in interaction with spectral tilt.

In all of the experiments, vowels were more likely to be identified as long if they had been produced with a voiced coda than if they had been produced with a voiceless coda. The stimuli in Experiments 1–3 were made from recordings of English speakers, so differences in vowel duration in the base recordings necessitated different manipulations for vowels from each environment, which could have contributed to the differences in perceived duration. However, Experiment 4 demonstrates that this difference is present even when stimuli are based on recordings of Telugu, a language that lacks significant voicing-conditioned vowel duration differences (Reddy, 1988; Sanker, 2018). Thus, the acoustic effects driving differences in perceived duration with English stimuli are not just an artifact of the manipulations required for vowels with different original durations.

Two main acoustic characteristics are likely to be contributing to differences in perceived duration based on the original coda. Intensity contour and spectral tilt are both influenced by coda voicing and also influence perceived vowel duration. Vowels before voiceless codas have a larger decrease in intensity than vowels before voiced codas (Archer et al., 2016; House & Fairbanks, 1953), and Experiment 2 demonstrates that falling intensity decreases perceived vowel duration. Spectral tilt is higher before non-glottalized voiceless codas than before voiced codas (Coleman, 2003; Sanker, 2019a), and Experiment 3 demonstrates that higher spectral tilt decreases perceived vowel duration. The acoustic measurements of the stimuli used in these experiments are consistent with differences in intensity contour and spectral tilt that have been reported previously as effects of coda voicing. All of the stimuli exhibited lower spectral tilt and higher final intensity before voiced codas than before voiceless codas, although some of the differences were small.

The results of all of these experiments demonstrate that coda voicing influences preceding vowels in ways that make them sound longer. Moreover, they demonstrate two acoustic characteristics that seem to be responsible for this effect: intensity contour and spectral tilt. These differences in perceived duration create a situation in which listeners could interpret perceived vowel duration as a veridical reflection of actual duration and develop distinct vowel durations conditioned by the coda voicing environment. Once vowels have different target durations in each environment, that duration difference may continue to increase, based on vowels still sounding longer than their actual duration in voiced environments, resulting in languages with substantial voicing-conditioned vowel duration. On the other hand, some languages might not phonologize the differences in perceived duration. If duration of the preceding vowel is not encoded as part of the bundle of acoustic traits that characterize coda voicing, both coda voicing environments are likely to retain similar vowel durations. This perceptual pathway for voicing-conditioned vowel duration does not exclude the possibility that articulatory factors also contribute to the effect, particularly when considering that even languages without large voicing-conditioned vowel duration differences often have small, inconsistent differences, as is observed for Arabic (de Jong & Zawaydeh, 2002; Port et al., 1980).

All of the experiments used English speakers as listeners, so it is not guaranteed that speakers of other languages would exhibit the same perceptual effects. However, this proposal for the development of voicing-conditioned vowel duration predicts that the acoustic influences of the original coda will influence listeners' perception of vowel duration similarly regardless of their native language. Effects of intensity contour and spectral tilt have non-linguistic parallels: Perception of duration in non-linguistic stimuli decreases with falling intensity (Grassi & Darwin, 2006; Schlauch et al., 2001) and increases with

higher overall intensity (Berglund et al., 1969; Goldstone & Lhamon, 1974). The existence of these non-linguistic parallels in perceptual effects of intensity are consistent with these effects not depending on an individual's native language. Thus, for example, speakers of Telugu would probably be influenced by the original coda voicing environment in the same way that English speakers are. On the other hand, compensation for the spliced ending is likely to be language-specific, because it depends on learned expectations about the relationship between vowel duration and coda voicing. If speakers are accustomed to vowels having similar durations in each coda voicing environment, the voicing of a spliced ending will not influence how long they expect the vowel to be.

5. Conclusions

These studies demonstrate several characteristics aside from duration itself that influence perceived vowel duration. Some effects suggest compensation for expected duration, while others do not seem to result from phonological expectations. The latter group of effects provides a possible pathway for the development of voicing-conditioned vowel duration.

Compensation for expected duration is reflected in two patterns: Splicing in a voiceless coda increases perceived duration, and high vowels have longer perceived duration than lower vowels do. Compensation for expected vowel duration in these conditions is consistent with vowel duration being part of the representation of coda voicing and vowel height in English, as previous work has also suggested.

Other effects are likely to reflect basic perceptual influences, rather than mapping onto an existing contrast or compensating for expected duration. The effects of spectral tilt and intensity can be explained by the greater perceived duration of vowels with greater perceived intensity and with rising intensity, respectively. Both effects of intensity have also been demonstrated with non-linguistic stimuli. Moreover, both are correlates of coda voicing, which could provide a perceptual pathway for the development of voicing-conditioned vowel duration. If these characteristics make vowels before voiced codas sound longer, listeners could interpret the perceived duration in each coda environment as reflecting the intended duration, resulting in longer vowels before voiced codas.

Additional Files

The additional files for this article can be found as follows:

- **Appendix.** A PDF document providing acoustic measurements for the stimuli, additional regression models, and additional figures. DOI: <https://doi.org/10.5334/labphon.268.s1>
- **Supplementary Material.** A zip file containing the raw response data for each of the four primary experiments and the follow-up to experiment 1, in CSV format. DOI: <https://doi.org/10.5334/labphon.268.s2>

Competing Interests

The author has no competing interests to declare.

References

- Abdelli-Beruh, N. B.** (2003). The stop voicing contrast in French sentences: Contextual sensitivity of vowel duration, closure duration, voice onset time, stop release and closure voicing. *Phonetica*, 61, 201–219. DOI: <https://doi.org/10.1159/000084158>
- Archer, S. L., Zamuner, T., Engel, K., Fais, L., & Curtin, S.** (2016). Infants' discrimination of consonants: Interplay between word position and acoustic saliency. *Language Learning and Development*, 12(1), 60–78. DOI: <https://doi.org/10.1080/15475441.2014.979490>

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J.** (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. DOI: <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S.** (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Bauer, M.** (2011). Articulatory conflict and laryngeal height. In *Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong, China* (pp. 292–295).
- 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. DOI: <https://doi.org/10.1006/jpho.2000.0128>
- Berglund, B., Berglund, U., Ekman, G., & Frankehaeuser, M.** (1969). The influence of auditory stimulus intensity on apparent duration. *Scandinavian Journal of Psychology*, 10(1), 21–26. DOI: <https://doi.org/10.1111/j.1467-9450.1969.tb00003.x>
- Casillas, J. V.** (2015). Production and perception of the /i/-/ɪ/ vowel contrast: The case of L2-dominant early learners of English. *Phonetica*, 72, 182–205. DOI: <https://doi.org/10.1159/000431101>
- Chen, M.** (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22, 129–159. DOI: <https://doi.org/10.1159/000259312>
- Chomsky, N., & Halle, M.** (1968). *The sound pattern of English*. Harper & Row.
- Chong, A. J., & Garellek, M.** (2018). Online perception of glottalized coda stops in American English. *Laboratory Phonology*, 9(1), Article 4. DOI: <https://doi.org/10.5334/labphon.70>
- Coleman, J.** (2003). Discovering the acoustic correlates of phonological contrasts. *Journal of Phonetics*, 31, 351–372. DOI: <https://doi.org/10.1016/j.wocn.2003.10.001>
- Cooper, W. E., & Danly, M.** (1981). Segmental and temporal aspects of utterance-final lengthening. *Phonetica*, 38, 106–115. DOI: <https://doi.org/10.1159/000260017>
- Coretta, S.** (2019). An exploratory study of voicing-related differences in vowel duration as compensatory temporal adjustment in Italian and Polish. *Glossa*, 4(1), Article 125. DOI: <https://doi.org/10.5334/gjgl.869>
- Crowther, C. S., & Mann, V.** (1992). Native language factors affecting use of vocalic cues to final consonant voicing in English. *Journal of the Acoustical Society of America*, 92(2), 711–722. DOI: <https://doi.org/10.1121/1.403996>
- de Jong, K.** (1991). An articulatory study of consonant-induced vowel duration changes in English. *Phonetica*, 48, 1–17. DOI: <https://doi.org/10.1159/000261868>
- de Jong, K.** (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics*, 32, 493–516. DOI: <https://doi.org/10.1016/j.wocn.2004.05.002>
- 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. DOI: <https://doi.org/10.1006/jpho.2001.0151>
- Durvasula, K., & Luo, Q.** (2014). Voicing, aspiration, and vowel duration in Hindi. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of meetings on acoustics*. Acoustical Society of America. (060009). DOI: <https://doi.org/10.1121/1.4895027>
- Fletcher, H., & Munson, W. A.** (1933). Loudness, its definition, measurement and calculation. *Bell System Technical Journal*, 12(4), 377–430. DOI: <https://doi.org/10.1002/j.1538-7305.1933.tb00403.x>
- Fourakis, M., & Iverson, G. K.** (1984). On the ‘incomplete neutralization’ of German final obstruents. *Phonetica*, 41, 140–149. DOI: <https://doi.org/10.1159/000261720>

- Fowler, C. A.** (1981). A relationship between coarticulation and compensatory shortening. *Phonetica*, 38, 35–50. DOI: <https://doi.org/10.1159/000260013>
- 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. DOI: [https://doi.org/10.1016/S0095-4470\(19\)30244-X](https://doi.org/10.1016/S0095-4470(19)30244-X)
- Goldstone, S., & Lhamon, W. T.** (1974). Studies of auditory-visual differences in human time judgment: 1. Sounds are judged longer than lights. *Perceptual and Motor Skills*, 39(1), 63–82. DOI: <https://doi.org/10.2466/pms.1974.39.1.63>
- Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G.** (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, 25(1), 1–42. DOI: <https://doi.org/10.1006/cogp.1993.1001>
- Grassi, M., & Darwin, C. J.** (2006). The subjective duration of ramped and damped sounds. *Perception & Psychophysics*, 68(8), 1382–1392. DOI: <https://doi.org/10.3758/BF03193737>
- Gussenhoven, C.** (2007). A vowel height split explained: Compensatory listening and speaker control. In J. Cole & J. I. Hualde (Eds.), *Laboratory phonology*, 9, 145–172. Mouton de Gruyter.
- 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. DOI: <https://doi.org/10.1121/1.2143736>
- Hillenbrand, J., Ingrisano, D. R., Smith, B. L., & Flege, J. E.** (1984). Perception of the voiced–voiceless contrast in syllable-final stops. *Journal of the Acoustical Society of America*, 76(1), 18–26. DOI: <https://doi.org/10.1121/1.391094>
- House, A. S.** (1961). On vowel duration in English. *Journal of the Acoustical Society of America*, 33(9), 1174–1178. DOI: <https://doi.org/10.1121/1.1908941>
- House, A. S., & Fairbanks, G.** (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, 25(1), 105–113. DOI: <https://doi.org/10.1121/1.1906982>
- Javkin, H. R.** (1976). The perceptual basis of vowel duration differences associated with the voiced/voiceless distinction. *Report of the Berkeley Phonology Laboratory*, 1, 78–92. DOI: <https://doi.org/10.1121/1.2002209>
- Keating, P. A.** (1979). *A phonetic study of voicing contrast in Polish* (Unpublished doctoral dissertation). Brown University.
- Kewley-Port, D., & Watson, C. S.** (1994). Formant-frequency discrimination for isolated English vowels. *Journal of the Acoustical Society of America*, 95(1), 485–496. DOI: <https://doi.org/10.1121/1.410024>
- Kim, D., & Clayards, M.** (2019). Individual differences in the link between perception and production and the mechanisms of phonetic imitation. *Language, Cognition and Neuroscience*, 34(6), 769–786. DOI: <https://doi.org/10.1080/23273798.2019.1582787>
- Kluender, K. R., Diehl, R. L., & Wright, B. A.** (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, 16, 153–169. DOI: [https://doi.org/10.1016/S0095-4470\(19\)30480-2](https://doi.org/10.1016/S0095-4470(19)30480-2)
- 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. DOI: <https://doi.org/10.1016/j.wocn.2012.07.002>
- Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R.** (2015). *lmerTest: Tests in linear mixed effects models* [Computer software manual]. Retrieved

- from <https://CRAN.R-project.org/package=lmerTest> (R package version 2.0-29). DOI: <https://doi.org/10.18637/jss.v082.i13>
- Laeufer, C.** (1992). Patterns of voicing-conditioned vowel duration in French and English. *Journal of Phonetics*, 20(4), 411–440. DOI: [https://doi.org/10.1016/S0095-4470\(19\)30648-5](https://doi.org/10.1016/S0095-4470(19)30648-5)
- Lisker, L.** (1986). “Voicing” in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, 29(1), 2–11. DOI: <https://doi.org/10.1177/002383098602900102>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D.** (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. DOI: <https://doi.org/10.1016/j.jml.2017.01.001>
- Mermelstein, P.** (1978). Difference limens for formant frequencies of steady-state and consonant-bound vowels. *Journal of the Acoustical Society of America*, 63(2), 572–580. DOI: <https://doi.org/10.1121/1.381756>
- Mitleb, F.** (1984). Voicing effect on vowel duration is not an absolute universal. *Journal of Phonetics*, 12, 23–27. DOI: [https://doi.org/10.1016/S0095-4470\(19\)30847-2](https://doi.org/10.1016/S0095-4470(19)30847-2)
- Moreton, E.** (2004). Realization of the English postvocalic [voice] contrast in F1 and F2. *Journal of Phonetics*, 32, 1–33. DOI: [https://doi.org/10.1016/S0095-4470\(03\)00004-4](https://doi.org/10.1016/S0095-4470(03)00004-4)
- Myers, S., & Hansen, B. B.** (2007). The origin of vowel length neutralization in final position: Evidence from Finnish speakers. *Natural Language & Linguistic Theory*, 25, 157–193. DOI: <https://doi.org/10.1007/s11049-006-0001-7>
- Öhman, S.** (1967). Peripheral motor commands in labial articulation. *Speech Transmission Laboratory Quarterly Progress Status Report*, 30–63.
- Penney, J., Cox, F., & Szakay, A.** (2018). Weighting of coda voicing cues: Glottalization and vowel duration. In *Proceedings of INTERSPEECH 2018: 19th annual conference of the International Speech Communication Association* (pp. 1422–1426). DOI: <https://doi.org/10.21437/Interspeech.2018-1677>
- Peterson, G. E., & Barney, H. L.** (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24(2), 175–184. DOI: <https://doi.org/10.1121/1.1906875>
- Peterson, G. E., & Lehiste, I.** (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32(6), 693–703. DOI: <https://doi.org/10.1121/1.1908183>
- Pierce, J. W.** (2007). PsychoPy – Psychophysics software in Python. *Journal of Neuroscience Methods*, 162, 8–13. DOI: <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Port, R. F., Al-Ani, S., & Maeda, S.** (1980). Temporal compensation and universal phonetics. *Phonetica*, 37, 235–252. DOI: <https://doi.org/10.1159/000259994>
- Port, R. F., & Dalby, J.** (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics*, 32(2), 141–152. DOI: <https://doi.org/10.3758/BF03204273>
- Port, R. F., & O'Dell, M. L.** (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics*, 13(4), 455–471. DOI: [https://doi.org/10.1016/S0095-4470\(19\)30797-1](https://doi.org/10.1016/S0095-4470(19)30797-1)
- Pycha, A., & 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. DOI: <https://doi.org/10.1016/j.wocn.2016.01.002>
- Reddy, K. N.** (1988). The duration of Telugu speech sounds: An acoustic study. *IETE Journal of Research*, 34(1), 57–63. DOI: <https://doi.org/10.1080/03772063.1988.11436705>

- Robinson, D. W., & Dadson, R. S.** (1956). A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics*, 7(5), 166–181. DOI: <https://doi.org/10.1088/0508-3443/7/5/302>
- Sanker, C.** (2018). Effects of laryngeal features on vowel duration: Implications for Winter's Law. *Papers in Historical Phonology*, 3, 180–205. DOI: <https://doi.org/10.2218/pihph.3.2018.2898>
- Sanker, C.** (2019a). Effects of coda voicing on vowel phonation. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 3323–3327).
- Sanker, C.** (2019b). Influence of coda stop features on perceived vowel duration. *Journal of Phonetics*, 75, 43–56. DOI: <https://doi.org/10.1016/j.wocn.2019.04.003>
- Schlauch, R. S., Ries, D. T., & DiGiovanni, J. J.** (2001). Duration discrimination and subjective duration for ramped and damped sounds. *Journal of the Acoustical Society of America*, 109(6), 2880–2887. DOI: <https://doi.org/10.1121/1.1372913>
- Seyfarth, S., & Garellek, M.** (2015). Coda glottalization in American English. In *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, Scotland 2015*.
- Seyfarth, S., & Garellek, M.** (2018). Plosive voicing acoustics and voice quality in Yerevan Armenian. *Journal of Phonetics*, 71, 425–450. DOI: <https://doi.org/10.1016/j.wocn.2018.09.001>
- Sharf, D. J.** (1964). Vowel duration in normal and whispered speech. *Language and Speech*, 7, 89–97. DOI: <https://doi.org/10.1177/002383096400700204>
- Solé, M.-J.** (2007). Controlled and mechanical properties in speech: A review of the literature. In M.-J. Solé, P. S. Beddor, & M. Ohala (Eds.), *Experimental approaches to phonology* (pp. 302–321). Oxford University Press.
- Solé, M.-J., & Ohala, J. J.** (2010). What is and what is not under the control of the speaker: Intrinsic vowel duration. In C. Fougeron, B. Kuehnert, M. Imperio, & N. Vallee (Eds.), *Laboratory phonology 10* (pp. 607–655). Mouton de Gruyter.
- 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. DOI: <https://doi.org/10.1121/1.395284>
- Summers, W. V.** (1988). F1 structure provides information for final-consonant voicing. *Journal of the Acoustical Society of America*, 84(2), 485–492. DOI: <https://doi.org/10.1121/1.396826>
- Toivonen, I., Blumenfeld, L., Gormley, A., Hoiting, L., Logan, J., Ramlakhan, N., & Stone, A.** (2015). Vowel height and duration. In U. Steindl et al. (Eds.), *Proceedings of the 32nd WCCFL* (pp. 64–71).
- Umeda, N.** (1975). Vowel duration in American English. *Journal of the Acoustical Society of America*, 58(2), 434–445. DOI: <https://doi.org/10.1121/1.380688>
- Wang, W. S.-Y., Lehiste, I., Chuang, C.-K., & Darnovsky, N.** (1976). Perception of vowel duration. *Journal of the Acoustical Society of America*, 60(S1), S92. DOI: <https://doi.org/10.1121/1.2003607>
- Warner, N., Jongman, A., Sereno, J., & Kamps, R.** (2004). Incomplete neutralization and other sub-phonemic duration differences in production and perception: Evidence from Dutch. *Journal of Phonetics*, 32, 251–276. DOI: [https://doi.org/10.1016/S0095-4470\(03\)00032-9](https://doi.org/10.1016/S0095-4470(03)00032-9)

How to cite this article: Sanker, C. 2020 A perceptual pathway for voicing-conditioned vowel duration. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1):18, pp.1–28. DOI: <https://doi.org/10.5334/labphon.268>

Submitted: 11 March 2020 **Accepted:** 23 September 2020 **Published:** 28 October 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

]u[*Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 