

Speaking rate, oro-laryngeal timing, and place of articulation effects on burst amplitude: Some cross-linguistic observations

Chandan R. Narayan

*Department of Languages, Literatures, and Linguistics
Speech and Psycholinguistics Lab
York University, Toronto*

Abstract

The relationship between speaking rate and burst amplitude was investigated in plosives with differing oro-laryngeal timing: long-lag voice-onset time (North American English) and short-lag voice-onset time (Indian Tamil). Burst amplitude (reflecting both intraoral pressure and flow geometry of the oral channel) was hypothesized to decrease in pre-vocalic plosives syllables with increasing speaking rate, which imposes temporal constraints on both intraoral pressure buildup behind the oral occlusion as well as respiratory air flow. Results showed that decreased vowel duration (which is associated with increased speaking rate) led to decreased burst amplitude in both short- and long-lag plosives. Aggregate models of bilabial and velar plosives (found both languages) suggested lower burst amplitudes in short-lag stops. Place-of-articulation effects in both languages were consistent with models of stop consonant acoustics, and place interactions with vowel duration were most apparent with long-lag English stops. Results are discussed in terms of speaking rate and language-internal forces contributing to burst amplitude variation and their implications for speech perception and potential to affect lenition phenomena.

1. Introduction

1.1. Background

Speaking rate is a highly variable phenomenon affected by a host of linguistic and extra-linguistic factors, resulting in adjustments to articulatory targets (Gay, 1981) and velocities (Adams, Weismar, and Kent, 1993; Dromey and Ramig, 1998). The phonetic consequences of modulations of speaking rate broadly affect both consonant temporal characteristics (e.g., voice-onset time, Kessinger and Blumstein, 1997) as well as spectral properties (e.g., vocalic trajectories, Lindblom, 1983). Recently, there has been considerable attention given to the phonetic effects of speaking rate in the context of (non-phonologized) consonant *lenition*, with much of the literature focused on intervocalic lenition processes in a variety of languages as indexed by consonant intensity (Warner and Tucker, 2011; Priva and Gleason, 2020) (see Lavoie, 2001 for a review) and consonant duration. The general pattern is for increased speaking rate (either implemented directly or as a consequence of style or register shifts) to result in increased consonant intensity (in VCVs) and reduced consonant duration (Soler and Romero, 1999; Warner and Tucker, 2011; Winter and Grawunder, 2012; Priva and Gleason, 2020). Lenition of this sort affects perception (Tucker, 2011) such that words with phonetically reduced variants are difficult to process. While a clear picture is emerging of how speaking rate modulates post- and inter-vocalic consonant amplitude and timing, very little is known about the relationship between speaking rate and the amplitude of initial prevocalic consonants, and in particular the amplitude characteristics of bursts, an acoustic feature that contributes to the perception of place-of-articulation (Ohde and Stevens, 1983; Repp and Lin, 1989).

The release of an oral stop consonant results in transient energy across the frequency spectrum, which is generally followed by broadband noise, collec-

tively taken as the “burst.” The transient portion of the burst, reflecting the release of pressurized intraoral air posterior the oral occlusion, together with the noise or aspiration following the transient, reveals the resonant properties of the tube anterior the oral occlusion being excited by air escaping the lungs prior to phonation, e.g., velars generally have a mid-frequency prominence as a result of a relatively long oral cavity. Burst amplitudes for voiceless stops are generally higher than for voiced stops (Stevens, Manuel, and Matthies, 1999; Chodroff and Wilson, 2014), reflecting greater intraoral pressure (P_o) and airflow at consonant release (Subtelny, Worth, and Sakuda, 1966; Haag, 1977). Both spectral envelope as well as amplitude characteristics at certain frequencies have been shown to be perceptually relevant for both place of articulation and voicing, respectively (Winitz, Scheib, and Reeds, 1972; Blumstein and Stevens, 1979; Ohde and Stevens, 1983; Repp and Lin, 1989; Krull, 1990; Chodroff and Wilson, 2014).

Although a buildup of P_o is necessary for the plosive burst characterizing oral stop consonants, it is subject to naturally occurring variation caused by extra-linguistic factors like stress and emotional state (Laukkanen et al., 1996) or affective voice quality such as whispering (Murry and Brown Jr, 1976). Models of intraoral pressure for voiceless stops suggest that the temporal constraints imposed by increased speaking rate need not necessarily affect P_o . (Müller and Brown Jr, 1980) (following Rothenberg, 1966) show P_o in voiceless stops has a rise time on the order of 50 ms (beginning before complete oral occlusion), which is potentially sufficient time to achieve P_o maxima in speeded speaking rate conditions where closure duration at fasts rates are generally greater than 50 ms (Allen and Miller, 1999; Pickett, Blumstein, and Burton, 1999). Nonetheless, some studies suggest that P_o is indeed correlated with speaking rate. In a study of Swedish aspirated “pa” spoken by a single speaker at different rates

(1-3 syllables/sec) (Hertegård, Gauffin, and Lindestad, 1995) found a high correlation between subglottal pressure and P_o , which decreased by approximately 4-5 cm H₂O at fast speaking rates. Miller and Daniloff (1977) examined P_o in three speakers' productions of English CVCs in monologues (thought to be comparable to conversational speaking rate) and citation contexts. For all three speakers, P_o for voiceless stops was at least 1 cm H₂O lower in monologue than citation contexts. Speaking rate also affects physiological properties of respiration. Lung volumes in speeded speaking tasks are lower than in slow speech (Dromey and Ramig, 1998), which in turn results in lower subglottal pressure (Sundberg, 2018) roughly corresponding to P_o during complete closure in plosive production.

Given the potential effect of speaking rate variation on P_o , comparable effects on burst amplitude are expected, however the extant literature on this relationship is mixed. In their study of P_o and speaking style, (Miller and Daniloff, 1977) also reported on peak absolute sound pressure level at consonant release, which showed mixed results between the monologue and citation consonant bursts. Burst amplitude was an absolute measure of the maximum sound pressure level in the r.m.s trace of release. One of the three speakers showed a correlated effect of speaking style on burst amplitude, which was lower in monologue speech than citation in all three places of articulation. The two other speakers exhibited varying effects of speaking style. The authors note that in their data, burst amplitude was correlated with both P_o and closure duration in two speakers, suggesting that it is not straightforwardly a reflection of P_o and that aerodynamic variables along with additional physiological parameters must be incorporated into models of normal speech. Indeed speakers' passive expansion of the oral cavity (e.g., depressing tongue body, expanding cheeks, etc.) and tissue compliance have been shown to slow the buildup of P_o during

stop closure, and are highly variable both within and across speakers, thereby contributing to the complex effects on burst amplitude (Bell-Berti, 1975; Westbury, 1983). These results contrast with the clear speech literature showing clear effects of style. Picheny, Durlach, and Braida (1986) report the effects of speaking style (clear speech versus conversational) on pre- and post-vocalic oral consonants in nonsense sentences as spoken by three American English speakers. Oral stops in their study consistently had more power in citation speech (5-10 dB) than in conversational speech. The results of Miller and Daniloff (1977) and Picheny, Durlach, and Braida (1986) suggest that burst amplitude in oral plosives is variable and affected by speaking style, which itself is correlated with speaking rate.

1.2. Research goals and predictions

The present research tests 1) the hypothesis that burst amplitude in CV syllables decreases with increasing speaking rate, and 2) whether such speaking rate effects are more or less evident at various places of articulation. This more general effect of the temporal constraints imposed by speaking rate may have been obscured in previous research by an absolute burst amplitude measure (Miller and Daniloff, 1977). In this study, we examine the effect of speaking rate on relative measures of burst amplitude, where peak amplitudes in various bands of burst noise are analyzed in relation to the amplitude of the following vowel. Such a normalization accounts for the overall amplitude of the utterance and may be a better index of the perceptual adjustments made by listeners when assessing place of articulation (Ohde and Stevens, 1983).

The present study also explores the relationship between burst amplitude, speaking rate, and oro-laryngeal synchronization, which is phonetically implemented in terms of short/long lag voice-onset time (VOT). It is hypothesized that burst amplitude is related to VOT, with long-lag VOT being associated

with high P_o and burst amplitude. The faster initiation of vocal fold oscillation in short-lag VOTs suggests that P_o is sufficiently low to promote a sufficient trans-glottal pressure differential for the rapid onset of phonation. This would be manifested in lower burst amplitude. The study focuses on consonants broadly classified as “voiceless,” that is, consonants where the primary acoustic and perceptual characteristic is voice onset time. Previous research on speaking rate, P_o , and burst amplitude was conducted in languages (English and Swedish) where pre-vocalic word-initial voiceless plosives are necessarily aspirated. The present study examines the relationship between relative burst amplitude and speaking rate in speakers of North American English (Indo-European) and Indian Tamil (Dravidian), a language which lacks phonemic voicing, and exhibits short-lag VOT word-initial plosives (Keane, 2004).

In comparing short-lag (Tamil) and long-lag (English) plosives, the present study addresses the contribution of speaking rate and oro-laryngeal timing dynamics to observed burst amplitude effects. Short-lag plosives from Tamil are examined instead of English short-lag stops (the voiced series /b d g/ [b̥ d̥ g̥]) in order to avoid aerodynamic and acoustic characteristics of phonologically voiced stops (such as closure voicing/voicing lead variability, passive cavity expansion during closure voicing, vocal fold adduction variability, F_0 perturbation) (Lindqvist, 1972; Flege, 1982; Lisker, 1986; Kingston and Diehl, 1994; Docherty et al., 2011) and associated P_o variability which might potentially obscure the effects of VOT on burst amplitude. Phonetic characteristics characteristic of phonological voicing in short-lag plosives in English have not been reported for short-lag Tamil plosives, where there is no such contrast. Given that the hypothesized mechanism affecting burst amplitude variation in the current study is P_o constrained by time in speeded conditions, the short-lag series from Tamil, rather than the phonologically English voiced series, was considered a more ap-

ropriate comparison with the long-lag English series. Language differences are therefore explained in terms of the effects of VOT type on burst amplitude.

Finally, the results of the study are discussed in the context of broader sources of variation, both language-internal (oro-laryngeal synchrony) and extra-linguistic (speaking rate), and directions for future study examining potential perceptual effects burst amplitude variation, are offered.

2. Methods

2.1. Speakers

The speech of twelve North American (Canadian and American) English speakers (six female and six male), aged 36-45 and ten Indian Tamil speakers (five female and five male), aged 24-44 was analyzed. All participants were native speakers of their respective languages, with no reported speech or hearing disorders. Participants were recruited via email solicitation and through various social media platforms.

2.2. Procedure

Due to restrictions to laboratory access during the global pandemic, participants were asked to record themselves in a quiet room at their respective residences (in Canada, America, and India). Participants were instructed to use a wired microphone (either ear-bud or closed-ear gaming style) along with recording software (such as Audacity or Praat) installed on their home computer. English-speaking participants were instructed to make three recordings, one each for the monosyllables “pa,” “ta,” and “ka.” For each recording, participants were asked to repeat the syllable at four self-paced speaking rates (slow, normal, fast, very fast) for approximately 5 seconds per rate. Instructions for Tamil-speaking participants were identical but with recordings for four places of articulation, bilabial ([pa]), dental ([ta]), velar ([ka]), and retroflex

([ta]). Instructions included an example audio file of the syllable “pa” (American English as recorded by the author) modeling the rate manipulation and providing a general structure of the recording. Participants made recordings with sampling rates at or above 22kHz and saved as **.wav** or **.aiff** files. Given the nature of the instructions (5 seconds per rate) participants produced more fast-rate tokens than slower rate tokens. There was some concern that intensity measurements would be affected by the variety of microphones and recording interfaces employed by the participants. No *post hoc* scaling was performed on raw amplitude measurements (burst) as difference measures (between the burst and adjacent vowel) served as dependent variables (see §2.3 below).

2.3. Measurements

All measurements were made by hand using Praat speech processing software by trained phoneticians. Two temporal measurements were made: duration of consonant release/burst (voice-onset time), and vowel duration. The onset of consonant release was identified at the first appearance of broadband transient noise, which was often followed by a burst of frication noise. Both transient and frication were considered the “burst.” Burst offset was identified at the zero-crossing before the first glottal cycle of the following vowel. The vowel onset (coinciding with the burst offset) was indexed by the first low-amplitude periodic oscillation. The end of the vowel (in the CV syllables) was marked by three co-occurring events: 1) a dramatic change in amplitude in the waveform, 2) a change in the energy in the formants accompanied by a change in complexity in the waveform indicating a loss of energy in F2 and F3, 3) the onset of aperiodicity.

Following Stevens, Manuel, and Matthies (1999), the spectrum of the burst was measured using an averaging technique with a 5ms Hamming window. The onset of periodic glottal oscillation of the vowel was avoided in the averaging

window. In cases where consonant release was immediately (<5ms) followed by glottal oscillation (generally in Tamil bilabial recordings), the unique measurement window was centered at the zero-crossing immediately preceding the onset of voicing. As one of our questions was motivated by potential place-specific effects of speaking rate induced burst amplitude variation, three amplitude measurements were taken reflecting important frequency bands for place perception following Ohde and Stevens (1983): 1) **AvF1**, or the average amplitude of F1 in the vowel, 2) **MaxHi**, or the maximum spectrum amplitude in the burst above 3kHz (for males) and 3.5kHz (for females), and 3) **MaxMid**, or the maximum spectrum amplitude in the burst in the F2/F3 range. Following Stevens, Manuel, and Matthies (1999), which identified two regions in the burst spectrum important for place perception, two relational measures were computed for each token, high-frequency **HiDiff** (AvF1-MaxHi) and mid-frequency **MidDiff** (AvF1-MaxMid) burst amplitudes. Relational measures of burst amplitude have been used other studies examining strength (e.g., DiCanio, 2012) and provides a token-specific normalization that mitigates the potential effect of varying Figure 1 shows the waveform and spectrogram representations of a typical token along with average spectra of the burst and vowel portions.

3. Results

3.1. Speaking rate and vowel duration

While some literature on speaking rate follows from a controlled implementation of speaking rate regulation procedures, where participants are instructed to synchronize their production with a metronome ensuring a consistent implementation of rate (e.g., Miller and Baer, 1983), much of the literature on rate effects proceed in a way similar to the present study, where participants vary their speaking speed in a subjective manner, e.g., Kessinger and Blumstein

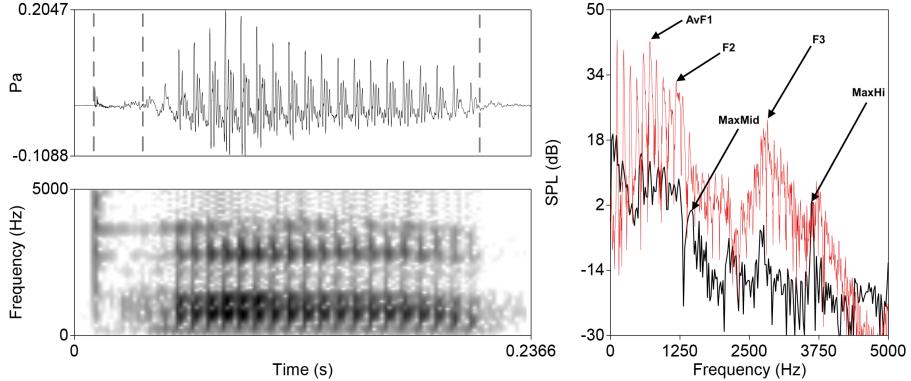


Figure 1: Waveform and wide-band spectrogram representations (left) of a typical “pa” token spoken by a male North American English speaker and averaged spectra for the burst and vowel (right). Three dashed lines on the waveform represent burst onset, burst offset/vowel onset, and vowel offset. Critical spectral measurements described in 2.3 are indicated on the burst spectrum (black) and vowel spectrum (red).

(1997) and Beckman et al. (2011). The modeling strategies in this literature rely on collapsing rate-dependent critical measurements (Gay, 1978; Kessinger and Blumstein, 1997; Soler and Romero, 1999). For example, dependent variables are analyzed as a function of a categorical independent rate variable with levels like “slow,” “fast,” etc. The present study is motivated by the question of acoustic consequences of durational shortening as induced by speaking rate, and as such, variability along the vowel duration *continuum*, treating it as a continuous variable. For this reason, vowel duration is taken as a proxy for subjective speaking rate, that is, it is assumed that faster speech results in shorter syllabic durations—but this may not necessarily be the case (e.g., speakers may vary the duration of inter-syllable intervals and thereby achieve slower or faster rates). In order to confirm this assumption, the relationship between subjective speaking rate and vowel duration was examined. The raw vowel duration data according to speaking rate and language is shown in Figure 2.

A linear mixed-effects model (`n1mle`) was fit to the vowel duration data with rate (slow, normal, fast, very fast) crossed with language (English, Tamil) as

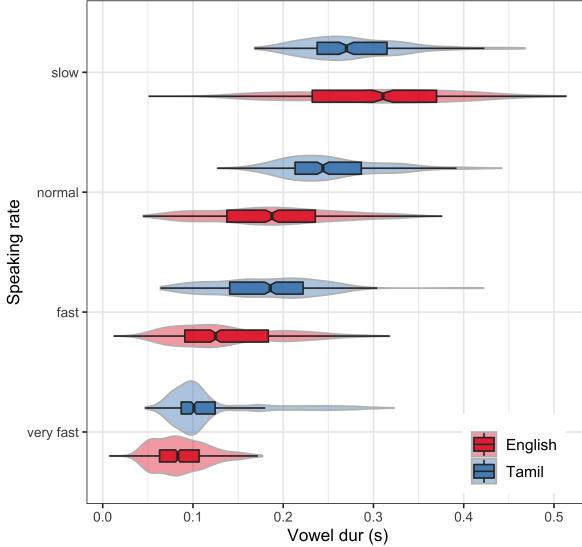


Figure 2: Observed vowel duration according to speaking rate and language. Violin plots show the density of vowel durations, while notched boxplots give medians and interquartile range.

predictors. The model was fit with random slopes and intercepts for rate by subject. Model coefficients are given in Table 1.

The model suggests that subjectively implemented speaking rate adjustments affect vowel duration in the direction predicted, increasing with decreasing subjective rate. The interaction between rate and language reflects the longer mean vowel duration for slow rate in English speakers than Tamil speakers, which flips the more general trend in the data, where vowel durations in Tamil are longer than in English in normal, fast, and very fast rates.

3.2. Vowel duration effects on burst amplitude

Vowel duration was used as a stand-in for speaking rate in all of the following analyses. Linear mixed-effects models were fit to examine 1) short-lag (Tamil) and long-lag (English) VOT differences in the effect of vowel duration on MidDiff and HiDiff across all places of articulation, 2) the language-specific interaction

Table 1: Mixed effects model of vowel duration as a function of subjective speaking rate and language.

Predictor	β	SE	t
(Intercept)	0.31	0.02	17.03
RateNormal	-0.11	0.01	-8.06
RateFast	-0.18	0.02	-8.94
RateVery fast	-0.23	0.02	-11.39
LanguageTamil	-0.03	0.03	-1.20
RateNormal \times LanguageTamil	0.09	0.02	4.13
RateFast \times LanguageTamil	0.09	0.03	2.81
RateVery fast \times LanguageTamil	0.07	0.03	2.29

Df(Rate effects)=3762; Df(Language)=19.

NB, References are RateSlow and LanguageEnglish.

Table 2: Mixed effects model of relative amplitude of stop bursts in A) the F2/F3 region (MidDiff) and B) above 3-3.5kHz (HiDiff), in English and Tamil.

Predictor	A: MidDiff			B: HiDiff		
	β	SE	t	β	SE	t
(Intercept)	32.48	3.31	9.82	36.71	2.93	12.55
Vowel dur	-33.81	8.88	-3.81	-29.27	7.15	-4.10
LangTamil	4.96	4.96	0.99	6.92	4.40	1.58
Vowel dur \times LangTamil	2.37	13.47	0.18	9.29	10.95	0.85

Df (Vowel dur)=3484; Df (Language)=20

NB, Reference level is LanguageEnglish.

effects of vowel duration and place of articulation on MidDiff and HiDiff. All models were fit with random slopes and intercepts for vowel duration by subject.

Model coefficients for both MidDiff and HiDiff are given in Table 2. The models show an overall negative effect of vowel duration on burst amplitude measures (MidDiff: $t=-3.81$, $p<0.001$; HiDiff: $t=-4.10$, $p<0.001$), with increasing vowel duration the difference in amplitude between the consonant and vowel decreases ($\sim 35\text{dB/sec}$). The models also show a small (though not significant) language difference with short-lag (Tamil) burst amplitudes being lower (MidDiff: $t=0.99$, $p=0.32$; HiDiff: $t=1.58$, $p=0.12$) than long-lag (English) burst amplitudes. Figure 3 shows the observed amplitude measurements as well as the model predictions for each language.

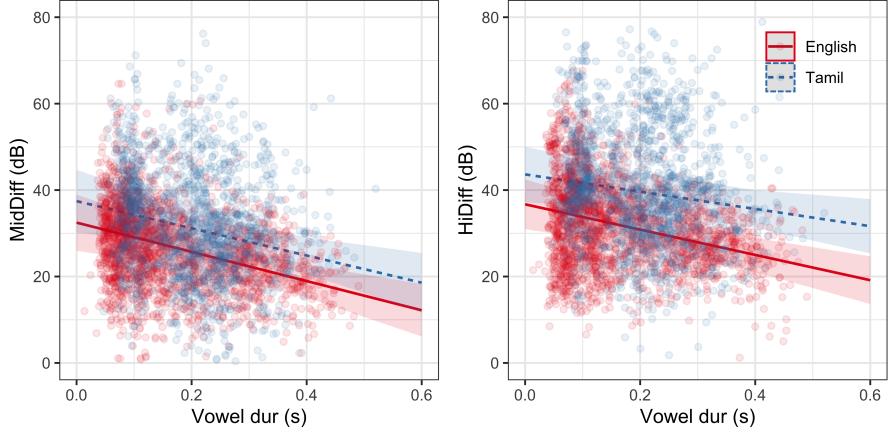


Figure 3: Predicted estimates (with confidence intervals) from linear mixed-effects models and observed values of MidDiff and HiDiff in long-lag (English) and short-lag (Tamil) plosives.

The results confirm the hypothesis that temporal constraints on articulation negatively affects the amplitude of release bursts in word-initial stops—with increasing vowel duration, the difference in amplitude between the consonant burst and the following vowel decreases, or put another way, with increasing vowel duration consonant bursts increase in amplitude. Model estimates also suggest that short-lag Tamil burst amplitude may be lower than in long-lag English plosives, especially for the high-frequency prominence at vowel durations longer than 200ms where confidence intervals no longer overlap. The language/VOT difference may be obscured by the differing consonant inventories in English and Tamil. To explore the possible language/VOT differences on burst amplitude, the next section models the effect of vowel duration on a subset of the data where English and Tamil share consonant places of articulation—bilabial and velar plosives.

3.3. Place-of-articulation and vowel duration effects on burst amplitude

In order to tease apart possible differences between short-lag (Tamil) and long-lag (English) plosives, a subset of the full data with only bilabial and

Table 3: Linear mixed effects models of relative amplitude of English and Tamil stop bursts in A) the F2/F3 region (MidDiff) and B) above 3-3.5kHz (HiDiff), as a function of vowel duration and consonant place of articulation.

Predictor	A: MidDiff			B: HiDiff		
	β	SE	t	β	SE	t
(Intercept)	36.67	2.64	15.03	46.98	2.35	20.03
LangTamil	8.12	4.01	2.02	6.89	3.61	1.91
POA-k	-16.60	0.67	-24.92	-11.93	0.69	-17.37
Vowel duration	-45.21	6.64	-6.80	-55.99	8.30	-6.74
LangTamil \times POA-k	-6.05	1.39	-4.37	-4.05	1.43	-2.83
LangTamil \times Vowel dur	15.85	10.48	1.51	43.36	12.90	3.36
POA-k \times Vowel dur	22.37	3.37	6.64	32.63	3.48	9.39
LangTamil \times POA-k \times Vowel dur	-10.75	6.51	-1.65	-27.73	6.71	-4.13

$Df_{POA, Vowel\ dur}=2064$. $Df_{Language}=20$. NB, References are POA-p and LangEnglish

velar stops was analyzed (shared by both Tamil and English). Fully crossed models (vowel duration \times POA \times language) of MidDiff and HiDiff were fit to the bilabial-velar data with random slopes and intercepts for vowel duration by subject. Table 3 gives model coefficients for both MidDiff and HiDiff.

The models coefficients suggest that short-lag (Tamil) burst amplitudes (for bilabials and velars) are lower (higher difference measures, β s \sim 7-8dB in both the mid- and high-frequency prominences) than in long-lag (English) plosives (MidDiff: $t=2.02$, $p=0.05$; HiDiff: $t=1.91$, $p=0.07$). The effect of vowel duration on burst amplitude is comparable to the models including all places of articulation in §3.2 (MidDiff: $t=-6.80$, $p<0.001$; HiDiff: $t=-6.74$, $p<0.001$).

Figure 4 gives the vowel duration varying model estimates of MidDiff and HiDiff for bilabials and velars in English. For the model of MidDiff burst amplitude, the relationship between bilabial and velar was different in Tamil relative to English ($t=-4.37$, $p<0.001$). The interaction between POA and vowel duration suggests that vowel duration has a different effect on velars than bilabials across both languages ($t=6.64$, $p<0.001$), and the three-way interaction term, although not significant ($t=-1.65$, $p=0.098$), suggests the effect of vowel du-

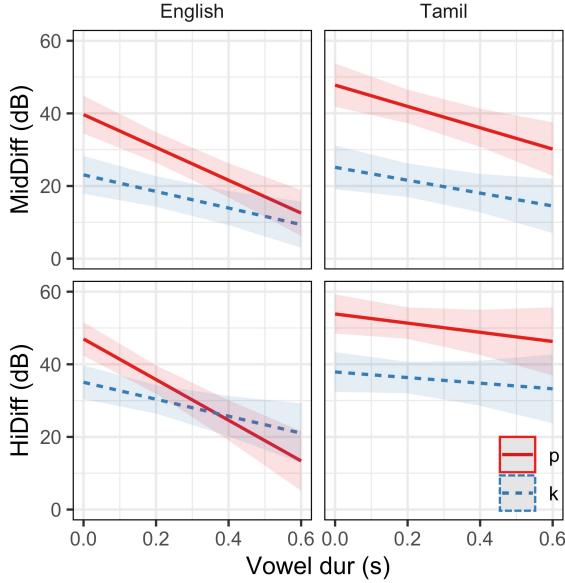


Figure 4: Predicted estimates (with confidence intervals) from linear mixed-effects models of MidDiff and HiDiff in bilabial and velar places in long-lag (English) and short-lag (Tamil) plosives.

ration on the relationship between bilabials and velars is different in the two languages (as evidenced by a slightly steeper slope for bilabials in English).

In the HiDiff model, the effect of vowel duration (across both POAs) on burst amplitude was different in short-lag (Tamil) relative to long-lag (English) plosives ($t=3.36, p<0.001$). The model also shows that the relationship between bilabial and velar was different in short-lag (Tamil) relative to long-lag (English) ($t=-2.83, p<0.005$). There was a significant interaction between POA and vowel duration suggesting that vowel duration has a different effect on velar burst amplitude than bilabial burst amplitude across both laryngeal specifications (English and Tamil) ($t=9.39, p<0.001$). Lastly, the three-way interaction term suggests differing effects of vowel duration on the relationship between bilabials and velars in the two languages/laryngeal specifications ($t=-4.13, p<0.001$).

Table 4: Linear mixed effects models of relative amplitude of English long-lag plosive bursts in A) the F2/F3 region (MidDiff) and B) above 3-3.5kHz (HiDiff), as a function of vowel duration and consonant place of articulation.

Predictor	A: MidDiff			B: HiDiff		
	β	SE	t	β	SE	t
(Intercept)	40.03	1.48	27.02	46.93	1.35	34.64
POA-t	-11.18	0.60	-18.71	-22.88	0.61	-37.77
POA-k	-16.60	0.62	-26.59	-12.43	0.63	-19.62
Vowel duration	-45.40	2.97	-15.27	-52.33	5.30	-9.87
POA-t \times Vowel dur	31.59	3.13	10.09	51.46	3.18	16.21
POA-k \times Vowel dur	21.97	3.15	6.96	33.82	3.21	10.55

Df = 2033. NB, Reference level is POA-p.

The effect of vowel duration on the various places of articulation in the two languages (with language-specific models) is explored in the next section.

3.4. Language-specific place-of-articulation effects on burst amplitude

Place-of-articulation effects in English and Tamil were analyzed separately to account for their differing stop-consonant inventories. While both languages share bilabial and velar places, coronal places vary—Tamil with two, dental ([t]) and retroflex ([ʈ]). Linear models of burst amplitude were fit with place of articulation and vowel duration as predictors and the same random-effects structure as the aggregate models above.

3.4.1. English long-lag plosives

The effects of vowel duration and place of articulation on MidDiff and HiDiff in English are given in Table 4.

All fixed effects and interactions were significant ($p < 0.001$) as were the vowel duration interactions with alveolar and velar places of articulation (relative to bilabials). That is, the effect of vowel duration on the burst amplitudes of alveolars and velars is different from vowel duration effects on the burst amplitudes of bilabials. These interactions are evident in the model estimates shown in Figure 5.

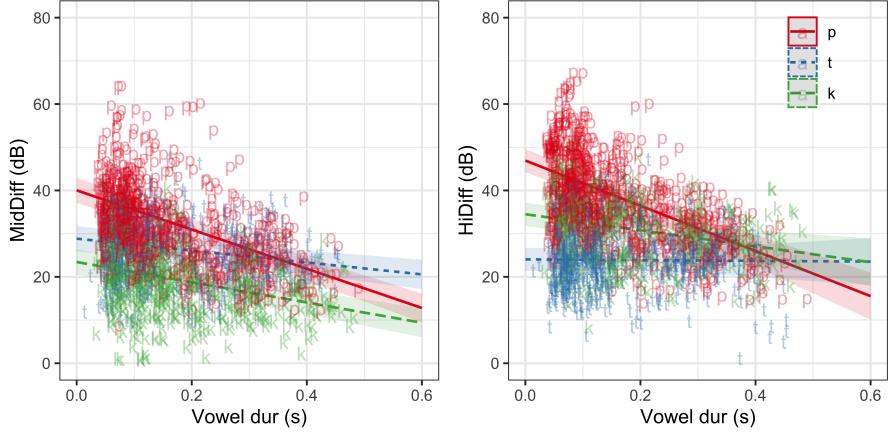


Figure 5: Predicted estimates (with confidence intervals) of MidDiff and HiDiff as a function of place of articulation and vowel duration in long-lag English plosives.

The overall effect of place of articulation on burst amplitude measures are consistent with those given in (Stevens, Manuel, and Matthies, 1999) and correspond to peak P_o given in (Subtelny, Worth, and Sakuda, 1966), with bilabials having the highest difference measures (lowest amplitude relative to the vowel) in both regions of the spectrum (MidDiff = 40.03dB; HiDiff = 46.93dB).

Velars have a high-amplitude mid-frequency prominence (MidDiff = 23.43dB) and a high-frequency prominence that is between bilabials and alveolars in amplitude (HiDiff = 34.5dB). Alveolars have a mid-frequency amplitude higher (MidDiff = 28.85dB) than bilabials, and a high-amplitude, high-frequency prominence (HiDiff = 24.05dB).

3.4.2. Tamil short-lag plosives

The effects of vowel duration and place of articulation on MidDiff and HiDiff in Tamil are given in Table 5.

All fixed POA effects were significantly different from the reference bilabial ($p < 0.001$). Similar to the long-lag English data, and consistent with Stevens, Manuel, and Matthies (1999), bilabials had the lowest burst amplitudes in both

Table 5: Linear mixed effects models of relative amplitude of Tamil short-lag plosives in A) the F2/F3 region (MidDiff) and B) above 3-3.5kHz (HiDiff), as a function of vowel duration and consonant place of articulation.

Predictor	A: MidDiff			B: HiDiff		
	β	SE	t	β	SE	t
(Intercept)	46.87	3.47	13.50	52.50	3.73	14.06
POA-t̄	-5.73	1.33	-4.32	-14.00	1.57	-8.94
POA-t̄	-16.13	1.41	-11.47	-13.70	1.59	-8.62
POA-k	-21.55	1.39	-15.55	-13.17	1.50	-8.78
Vowel duration	-23.01	9.73	-2.36	-8.21	11.64	-0.70
POA-t̄ × Vowel dur	-9.93	6.13	-1.62	-3.11	7.17	-0.43
POA-t̄ × Vowel dur	1.12	6.48	0.17	-16.57	7.32	-2.26
POA-k × Vowel dur	5.60	6.35	0.88	2.61	6.93	0.38

Df = 1441. NB, Reference level is POA-p.

mid- and high-frequency ranges (MidDiff = 46.87dB; HiDiff = 52.50dB). Velars had the highest burst amplitude in the mid-frequency range, (MidDiff = 25.32dB) and retroflexes followed by dentals had the highest burst amplitude in the high-frequency range (HiDiff_{ret} = 38.80dB, HiDiff_{den} = 38.50dB), again consistent with Stevens, Manuel, and Matthies (1999) which showed alveolars with the highest amplitude in that range.

Figure 6 shows the model estimates of the burst amplitudes for each place of articulation as a function of vowel duration. The vowel duration effects on the mid-frequency burst amplitudes were similar at all places of articulation (i.e., there are no significant interactions). While the overall effect of vowel duration on high-frequency burst amplitudes was not significant ($t=-0.70$, $p=0.48$), retroflex stops showed a significantly steeper increase in amplitude with increasing vowel duration relative to bilabials ($t=-2.26$, $p<0.05$).

3.5. Voice onset time, vowel duration and burst amplitude in English and Tamil

To better understand the oro-laryngeal timing differences in burst amplitude as a function of speaking rate, the relationship between VOT and vowel duration

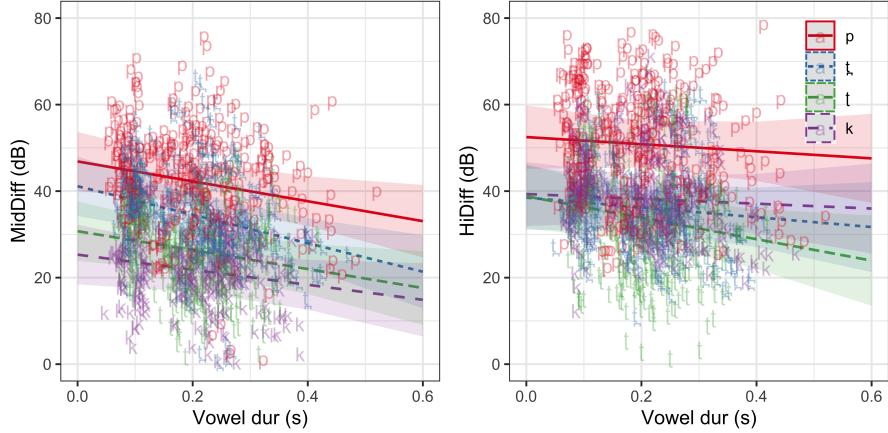


Figure 6: Predicted estimates (with confidence intervals) and observed values of MidDiff and HiDiff as a function of place of articulation and vowel duration in short-lag Tamil plosives.

Table 6: Mixed effects model of VOT as a function of vowel duration in long-lag English and short-lag Tamil plosives.

Predictor	β	SE	t
(Intercept)	0.01	0.002	4.35
Vowel duration	0.19	0.02	9.19
Language _{Tamil}	-0.001	0.004	-0.22
Vowel dur \times Language _{Tamil}	-0.15	0.03	-4.75

$Df_{\text{Vowel dur}} = 3484$, $Df_{\text{Language}} = 20$; Df (Language) = 20. NB, Reference level is Language_{English}.

was modeled in the two languages. The results of the VOT model are given in Table 6 and estimates visualized in Figure 7.

Vowel duration has a clear positive effect on VOT ($t=9.19$, $p<0.001$). This result was consistent with literature showing that as speaking rate decreases, vowel duration and VOT increase in equal proportions (e.g., Kessinger and Blumstein, 1997; Theodore, Miller, and DeSteno, 2009), as well as the *laryngeal realism* literature suggesting that phonetic cues to phonological categories decrease in duration as speech rate increases (Beckman et al., 2011). The effect of vowel duration on VOT is different between the two languages, with short-lag Tamil plosives having a shallower slope ($t=-4.75$, $p<0.001$).

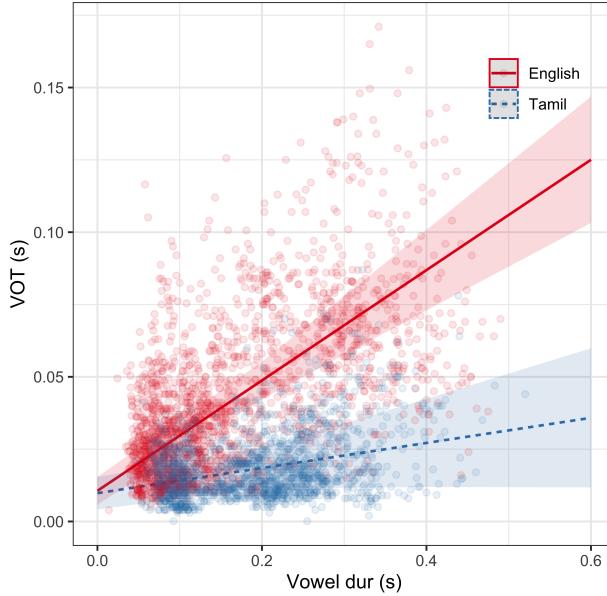


Figure 7: Predicted estimates (with confidence intervals) mixed-effects models and observed values of VOT in A) long-lag English and B) short-lag Tamil plosives.

Although there is an overall positive association between speaking rate and VOT, the size of the effect is reduced in Tamil short-lag implementation of VOT (75% of VOTs are less than 0.025s), while the long-lag VOT status in English allows for a more flexible oro-laryngeal timing. Given the relative immutability of VOT in short-lag Tamil plosives, does it contribute to burst amplitude variability? The effect of VOT on burst amplitude measures were modeled separately in both languages.

3.6. VOT effects on burst amplitude

3.6.1. Long-lag English plosives

Mixed effects models of MidDiff and HiDiff were fit to the data with VOT as a predictor and random intercepts and slopes for VOT by subject. Model coefficients are shown in Table 7.

Table 7: Linear mixed effects models of relative amplitude of long-lag English plosive bursts in A) the F2/F3 region (MidDiff) and B) above 3-3.5kHz (HiDiff), as a function of voice onset time.

Predictor	A: MidDiff			B: HiDiff		
	β	SE	t	β	SE	t
(Intercept)	34.47	1.59	21.68	38.77	2.08	18.57
VOT	-192.92	21.12	-9.13	-172.40	31.31	-5.51

Df (both models) = 2037.

Table 8: Linear mixed effects models of relative amplitude of Tamil short-lag plosives in A) the F2/F3 region (MidDiff) and B) above 3-3.5kHz (HiDiff), as a function of voice onset time.

Predictor	A: MidDiff			B: HiDiff		
	β	SE	t	β	SE	t
(Intercept)	36.78	3.81	9.64	43.13	4.17	10.33
VOT	-344.95	100.57	-3.43	-185.01	101.13	-1.83

Df (both models) = 1447.

Results show that VOT has a significant effect ($p < 0.001$) on burst amplitude in both the mid-frequency and high-frequency prominences. Similar to vowel duration, with increasing VOT, burst amplitude increases (difference measures decrease).

3.6.2. Short-lag Tamil plosives

Mixed models were likewise fit to the Tamil short-lag burst amplitude data, with VOT as a predictor and random slopes and intercepts for VOT by subject. Table 8 shows the model coefficients, which suggest a significant positive effect of VOT on burst amplitude in only the mid-frequency prominence ($t = -3.43$, $p < 0.001$), while the effect is muted ($t = -1.83$, $p < 0.1$) in the high-frequency prominence.

The overall temporal constraining of the Tamil syllable (by increasing speaking rate, which minimally decreases VOT) primarily affects the mid-frequency burst amplitude. That is, the effects of VOT are comparable to the effects of vowel duration on Tamil short-lag burst amplitudes (§3.4).

4. General discussion

Linguistic factors such as syllable structure (Mackay, 1974) and phrasal length (Yuan, Liberman, and Cieri, 2006), and extra-linguistic factors like emotional content of the speech (Mozziconacci and Hermes, 2000), and age and geographical background (Quené, 2005) of the speaker all have an effect on speaking rate. The current study investigated whether speaking rate variation imposes constraints on the articulation (and consequently aerodynamics and acoustics) of consonants via changes in burst amplitude, a proxy for intraoral pressure and glottal friction. Speaking rate (as evidenced in vowel duration) affected burst amplitude, such that decreasing vowel duration led to a decrease in amplitude (relative to the following vowel) in speakers' consonant bursts. While the effect is found in both long-lag (North American English) and short-lag (Indian Tamil) plosives, it is more pronounced in the long-lag series, as revealed by its interaction with place of articulation.

Results also suggested that burst amplitudes of bilabials and velars (found in both languages) were lower in short-lag (Tamil) than in long-lag (English) plosives. These differences were unpacked by examining the association between vowel duration and VOT. Consistent with the extant literature, the long-lag VOT is more tightly associated with vowel duration. While there is an effect of vowel duration on VOT in short-lag plosives, it is considerably more shallow. This suggests that the differing oro-laryngeal timing in the two types of plosives may be responsible for the burst amplitude variation. Long-lag VOTs are strongly correlated with vowel duration as well as higher overall burst amplitudes than in short-lag plosives. When coupled with the general tendency for voiceless aspirated stops to have long closure durations (Lisker, 1957; Stathopoulos and Weismer, 1983), it can be deduced that the high burst amplitudes results from sufficient time for P_o buildup. Conversely, short-lag stops have lower burst

amplitudes than English (suggesting lower P_o), with less pronounced effects of speaking rate. However, the automatic relationship between short-lag VOT and burst-amplitude need not necessarily follow in a way exemplified in the present data. For example, Korean fortis stops show high P_o and have burst energy comparable to lenis stops despite having short-lag VOTs (Cho, Jun, and Ladefoged, 2002). This suggests that the link between short-lag VOT, low P_o and low burst amplitude may be an articulatory-aerodynamic default when there is no phonological laryngeal contrast for stops in initial position.

4.1. Conclusion and future directions

We can conclude from this study that speaking rate, as manifested in vowel duration, affects the burst amplitude of preceding consonants, and that VOT serves to mitigate that effect. These results are very much in line with a growing phonetics literature focused on rate variability and its contribution to phonological lenition (Kirchner, 2004; Warner and Tucker, 2011; Priva and Gleason, 2020), which has identified fast speech (in environments such as intervocalic and postvocalic positions) as affecting shorter consonantal duration and increased intensity. While the outcome of fast speech in this literature affects a non-phonologized synchronic lenition in production, directions for future research building upon the present results might consider how the acoustic variability (resulting from either speaking rate or language-specific VOT) in consonant bursts contribute to both our understanding of (1) general consonant perception, and (2) listener-driven diachronic patterns in languages.

From a purely speech processing perspective, we might ask about the extent to which the phonetic association between vowel duration and burst amplitude is represented in the listener, that is, does a listener expect an association between vowel duration (rate-conditioned or otherwise) and reduced burst amplitude? If so, how does a listener's language experience affect their perception of this

trading relationship (Repp, 1982)? Would the association between rate/vowel duration and burst amplitude be stronger long-lag than in short-lag plosives? Does the relationship serve as a perceptual cue for phonological voicing in English speakers (or other languages that contrast long- and short-lag VOT)? This line of research will contribute to not only our understanding of cue integration in consonant perception but also theories that center the role of experience with speech in individuals' perceptual capacities.

A second line of inquiry considers the laryngeal-timing and place effects observed in the present study as contributing to our understanding of diachronic lenition processes. Although burst characteristics are part of a larger constellation of acoustic cues utilized by the listener in determining consonant identity, disruption of burst cues results in a greater reliance on formant transitional cues in perception (Dorman, Studdert-Kennedy, and Raphael, 1977). When coupled with the tendency for speaking rate induced adjustments to formant transition patterns (Gay, 1978; Krull, 1989; Duez, 1992) we offer the possibility that low burst amplitude resulting from either language-internal forces or external factors like speaking rate, may potentially be misperceived by the listener (Ohala et al., 1981). Are listeners more or less sensitive to amplitude changes in different frequency bands (i.e., is a reduction in mid-frequency noise as observed in the current short-lag data significant for the listener) at different places of articulation? Do patterns in reduced burst amplitude perception, especially in naturally low-amplitude voiceless short-lag bilabials, resemble the types of historical lenition changes found in languages (e.g., debuccalization of bilabials in languages like Old Kannada and Japanese) (Gai, 1946; Sasaki, 2008)?

In this way, the study of naturally conditioned variation in stop consonant burst amplitude may allow us to better understand the nature of listener knowledge as well as how that knowledge potentially affects historical phenomena.

References

- Adams, Scott G, Gary Weismser, and Raymond D Kent (1993). “Speaking rate and speech movement velocity profiles”. *Journal of Speech, Language, and Hearing Research* 36.1, pp. 41–54.
- Allen, J Sean and Joanne L Miller (1999). “Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words”. *The Journal of the Acoustical Society of America* 106.4, pp. 2031–2039.
- Beckman, Jill et al. (2011). “Rate effects on Swedish VOT: Evidence for phonological overspecification”. *Journal of Phonetics* 39.1, pp. 39–49.
- Bell-Berti, Fredericka (1975). “Control of pharyngeal cavity size for English voiced and voiceless stops”. *The Journal of the Acoustical Society of America* 57.2, pp. 456–461.
- Blumstein, Sheila E and Kenneth N Stevens (1979). “Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants”. *The Journal of the Acoustical Society of America* 66.4, pp. 1001–1017.
- Cho, Taehong, Sun-Ah Jun, and Peter Ladefoged (2002). “Acoustic and aerodynamic correlates of Korean stops and fricatives”. *Journal of Phonetics* 30.2, pp. 193–228.
- Chodroff, Eleanor and Colin Wilson (2014). “Burst spectrum as a cue for the stop voicing contrast in American English”. *The Journal of the Acoustical Society of America* 136.5, pp. 2762–2772.
- DiCanio, Christian T (2012). “The phonetics of fortis and lenis consonants in Itunyoso Trique”. *International Journal of American Linguistics* 78.2, pp. 239–272.
- Docherty, Gerard et al. (2011). “Variation in Voice Onset Time along the Scottish-English Border.” In: *ICPhS*, pp. 591–594.

- Dorman, Michael F, Michael Studdert-Kennedy, and Lawrence J Raphael (1977). “Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues”. *Perception & Psychophysics* 22.2, pp. 109–122.
- Dromey, Christopher and Lorraine Olson Ramig (1998). “The effect of lung volume on selected phonatory and articulatory variables”. *Journal of Speech, Language, and Hearing Research* 41.3, pp. 491–502.
- Duez, Danielle (1992). “Second formant locus-nucleus patterns: an investigation of spontaneous French speech”. *Speech Communication* 11.4-5, pp. 417–427.
- Flege, James Emil (1982). “Laryngeal timing and phonation onset in utterance-initial English stops”. *Journal of Phonetics* 10.2, pp. 177–192.
- Gai, Govind Swamirao (1946). *Historical Grammar of Old Kannada: Based Entirely on the Kannada Inscriptions of the 8th, 9th and 10th Centuries AD*. Vol. 1. Deccan College, Postgraduate and Research Institute.
- Gay, Thomas (1978). “Effect of speaking rate on vowel formant movements”. *The Journal of the Acoustical Society of America* 63.1, pp. 223–230.
- (1981). “Mechanisms in the control of speech rate”. *Phonetica* 38.1-3, pp. 148–158.
- Haag, Winfried K (1977). “Oral Air Pressure and Airflow in German Stop Consonants”. *Phonetica* 34.1, pp. 31–39.
- Hertegård, Stellan, Jan Gauffin, and Per-Åke Lindestad (1995). “A comparison of subglottal and intraoral pressure measurements during phonation”. *Journal of Voice* 9.2, pp. 149–155.
- Keane, Elinor (2004). “Tamil”. *Journal of the International Phonetic Association* 34.1, pp. 111–116.

- Kessinger, Rachel H and Sheila E Blumstein (1997). “Effects of speaking rate on voice-onset time in Thai, French, and English”. *Journal of Phonetics* 25.2, pp. 143–168.
- Kingston, John and Randy L Diehl (1994). “Phonetic knowledge”. *Language* 70.3, pp. 419–454.
- Kirchner, Robert (2004). “Consonant lenition”. In: *Phonetically Based Phonology*. Ed. by Bruce Hayes, Robert Kirchner, and Donca Steriade. Cambridge: Cambridge University Press. Chap. 10, p. 313.
- Krull, Diana (1989). “Second formant locus patterns and consonant-vowel coarticulation in spontaneous speech”. *Phonetic Experimental Research at the Institute of Linguistics, University of Stockholm* 10, pp. 87–108.
- (1990). “Relating acoustic properties to perceptual responses: A study of Swedish voiced stops”. *The Journal of the Acoustical Society of America* 88.6, pp. 2557–2570.
- Laukkanen, Anne-Maria et al. (1996). “Physical variations related to stress and emotional state: a preliminary study”. *Journal of Phonetics* 24.3, pp. 313–335.
- Lavoie, Lisa M (2001). *Consonant strength: Phonological patterns and phonetic manifestations*. Routledge.
- Lindblom, Björn (1983). “Economy of speech gestures”. In: *The Production of Speech*. Ed. by Peter MacNeilage. Springer, pp. 217–245.
- Lindqvist, J (1972). *Laryngeal articulation studied on Swedish subjects*. Tech. rep. Royal Institute of Technology, Speech Transmission Laboratory.
- Lisker, Leigh (1957). “Closure duration and the intervocalic voiced-voiceless distinction in English”. *Language* 33.1, pp. 42–49.
- (1986). ““Voicing” in English: A catalogue of acoustic features signaling/b/versus/p/in trochees”. *Language and speech* 29.1, pp. 3–11.

- Mackay, Donald G (1974). "Aspects of the syntax of behavior: Syllable structure and speech rate". *Quarterly Journal of Experimental Psychology* 26.4, pp. 642–657.
- Miller, Creighton J and Ray Daniloff (1977). "Aerodynamics of stops in continuous speech". *Journal of Phonetics* 5.4, pp. 351–360.
- Miller, Joanne L and Thomas Baer (1983). "Some effects of speaking rate on the production of/b/and/w/". *The Journal of the Acoustical Society of America* 73.5, pp. 1751–1755.
- Mozziconacci, Sylvie JL and Dik J Hermes (2000). "Expression of emotion and attitude through temporal speech variations". In: *Sixth International Conference on Spoken Language Processing*.
- Müller, Eric M and William Samuel Brown Jr (1980). "Variations in the supraglottal air pressure waveform and their articulatory interpretation". In: *Speech and Language*. Vol. 4. Elsevier, pp. 317–389.
- Murry, Thomas and WS Brown Jr (1976). "Peak intraoral air pressures in whispered stop consonants". *Journal of Phonetics* 4.3, pp. 183–187.
- Ohala, John J et al. (1981). "The listener as a source of sound change". *Parasesision on language and behavior (Chicago Linguistics Society, Chicago 1981)*.
- Ohde, Ralph N and Kenneth N Stevens (1983). "Effect of burst amplitude on the perception of stop consonant place of articulation". *The Journal of the Acoustical Society of America* 74.3, pp. 706–714.
- Picheny, M., N. Durlach, and L. Braida (1986). "Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech." *Journal of Speech and Hearing Research* 29 4, pp. 434–46.
- Pickett, Emily R, Sheila E Blumstein, and Martha W Burton (1999). "Effects of speaking rate on the singleton/geminate consonant contrast in Italian". *Phonetica* 56.3-4, pp. 135–157.

- Priva, Uriel Cohen and Emily Gleason (2020). “The causal structure of lenition: A case for the causal precedence of durational shortening”. *Language* 96.2, pp. 413–448.
- Quené, Hugo (2005). “Modeling of between-speaker and within-speaker variation in spontaneous speech tempo”. In: *Ninth European Conference on Speech Communication and Technology*.
- Repp, Bruno H (1982). “Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception.” *Psychological Bulletin* 92.1, p. 81.
- Repp, Bruno H and Hwei-Bing Lin (1989). “Acoustic properties and perception of stop consonant release transients”. *The Journal of the Acoustical Society of America* 85.1, pp. 379–396.
- Rothenberg, Martin (1966). “The Breath-Stream Dynamics of Simple-Released-plosive Production”. PhD thesis. University of Michigan.
- Sasaki, Kan (2008). “Hardening alternation in the Mitsukaido dialect of Japanese”. *Gengo Kenkyu* 134, pp. 85–118.
- Soler, Antonia and Joaquín Romero (1999). “The role of duration in stop lenition in Spanish”. In: *Proceedings of the 14th International Congress of Phonetic Sciences*. Vol. 1. The Regents of the University of California Oakland, CA, pp. 483–486.
- Stathopoulos, Elaine T and Gary Weismer (1983). “Closure duration of stop consonants”. *Journal of Phonetics* 11.4, pp. 395–400.
- Stevens, Kenneth N, Sharon Y Manuel, and Melanie Matthies (1999). “Revisiting place of articulation measures for stop consonants: Implications for models of consonant production”. In: *Proceedings of the International Congress of Phonetic Sciences*, pp. 1117–1120.

- Subtelny, Joanne D, Joseph H Worth, and Mamoru Sakuda (1966). “Intraoral pressure and rate of flow during speech”. *Journal of Speech and Hearing Research* 9.4, pp. 498–518.
- Sundberg, Johan (2018). “The singing voice”. *The Oxford handbook of voice perception*, p. 117.
- Theodore, Rachel M, Joanne L Miller, and David DeSteno (2009). “Individual talker differences in voice-onset-time: Contextual influences”. *The Journal of the Acoustical Society of America* 125.6, pp. 3974–3982.
- Tucker, Benjamin V (2011). “The effect of reduction on the processing of flaps and/g/in isolated words”. *Journal of Phonetics* 39.3, pp. 312–318.
- Warner, Natasha and Benjamin V Tucker (2011). “Phonetic variability of stops and flaps in spontaneous and careful speech”. *The Journal of the Acoustical Society of America* 130.3, pp. 1606–1617.
- Westbury, John R (1983). “Enlargement of the supraglottal cavity and its relation to stop consonant voicing”. *The Journal of the Acoustical Society of America* 73.4, pp. 1322–1336.
- Winitz, Harris, Marlin E Scheib, and James A Reeds (1972). “Identification of stops and vowels for the burst portion of/p, t, k/isolated from conversational speech”. *The Journal of the Acoustical Society of America* 51.4B, pp. 1309–1317.
- Winter, Bodo and Sven Grawunder (2012). “The phonetic profile of Korean formal and informal speech registers”. *Journal of Phonetics* 40.6, pp. 808–815.
- Yuan, Jiahong, Mark Liberman, and Christopher Cieri (2006). “Towards an integrated understanding of speaking rate in conversation”. In: *Ninth International Conference on Spoken Language Processing*.