Speaking rate and language-specific voice onset time effects on burst amplitude: Cross-linguistic observations and implications for sound change

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 in voiceless plosives was investigated in two languages with differing oro-laryngeal timing implementations of phonological voicing, North American English and Indian Tamil. Burst amplitude (reflecting both intraoral pressure and flow geometry of the oral channel) was hypothesized to decrease in CV syllables with increasing speaking rate, which imposes temporal constraints on both intraoral pressure buildup behind the oral occlusion as well as respiratory air flow. Increased speaking rate led to decreased burst amplitude (relative to vowel amplitude) in both languages, with the magnitude of the effect being considerably weaker in Tamil, which has short-lag implementation of voice onset time. Bilabials in both languages were affected disproportionately relative to other places of articulation. Additionally, burst amplitudes were lower overall in Tamil, reflecting lower intraoral pressure which promotes faster vocalic onset. Results are discussed in terms of language-internal and extra-linguistic phonetic phenomena potentially serving as perceptual triggers for historical sound change.

---

1. Introduction

1.1. Purpose and 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

Preprint submitted to Elsevier *July 22, 2021*

velocities (Adams et al., 1993) with consequences for intelligibility (Krause & Braida, 2002). While research on the acoustic-phonetic effects of speaking rate have focused on time- and spectral- domain characteristics such as voice-onset time (Kessinger & Blumstein, 1997) and formant undershoot (Lindblom, 1983), little is known about the relationship between speaking rate and the amplitude of consonants, and in particular the amplitude characteristics of stop bursts, an acoustic feature that contributes to place of articulation perception (Ohde & Stevens, 1983).

The release of an oral stop consonant results in transient energy across the frequency spectrum, which is generally followed by broadband noise, collectively 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, reveal 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 et al., 1999; Chodroff & Wilson, 2014), reflecting greater intraoral pressure (P_o) and airflow at consonant release (Subtelny et al., 1966; Haag, 1977). Both spectral envelope as well as amplitude characteristics at certain frequencies have been shown to be perceptually relevant for place of articulation and voicing, respectively (Blumstein & Stevens, 1979; Ohde & Stevens, 1983; Chodroff & Wilson, 2014).

Although a buildup of P_o is required for the plosive effect 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 & Brown Jr, 1976). Models of intraoral pressure for voiceless stops suggest that the temporal cosntraints imposed by increased speaking rate need not necessarily affect P_o. Müller & 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 > 50 ms (Allen & Miller, 1999; Pickett et al., 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 et al. (1995) found a high correlation between subglottal pressure and P_o, which decreased by approximately 4-5 cm H_2O at fast speaking rates. Miller & 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 H_2O 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 & 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 & 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 et al. (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 & Daniloff (1977) and Picheny et al. (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

The current research directly tests the hypothesis that burst amplitude in CV syllables, being a function of P_o and pre-phonation glottal flow, decreases with increasing speaking rate. 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 & Daniloff, 1977). In this study, we examine the effect of speaking rate on relative measures of burst amplitude offered by (Stevens et al., 1999), 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 & Stevens, 1983).

The current study also explores the relationship between burst amplitude, speaking rate, and oro-laryngeal timing, which is phonologically specified in terms of short/long lag voice-onset time (VOT). It is hypothesized that burst amplitude, reflecting P_o and pre-phonation glottal aspiration noise, is related to VOT, with long-lag VOT being associated with high P_o.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 rapid phonation. 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 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) VOT languages, the current study addresses the contribution of speaking rate and phonologically determined oro-laryngeal to observed burst amplitude effects. Finally, the results of the current study are discussed in the context of perceptual explanations for sound change triggered by both language-internal and extra-linguistic phonetic phenomenoa.

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 (பா [pa]), dental (தா [t̪a]), velar (கா [ka]), and retroflex (டா [ʈa]). 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. Participants’ recordings indicated that successful speech rate manipulation was achieved, providing a wide range of syllable durations. Given the nature of the instructions (5 seconds per rate) participants produced more fast-rate tokens than slower rate tokens.

2.3. Measurements

All measurements were made by hand using Praat speech processing software by trained phoneticians. Two durational measuments 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 et al. (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. Three amplitude measurements were taken, again following Stevens et al. (1999): 1) **Av**, or the average amplitude of F1 in the vowel, 2) **Ahi**, or the maximum spectrum amplitude in the burst above 3kHz (for males) and 3.5kHz (for females), and 3) **max23**, or the maximum spectrum amplitude in the burst in the F2/F3 range. Two relational measures were computed for each token, high-frequency **HiDiff** (Av-Ahi) and mid-frequency **F23Diff** (Av-max23) burst amplitudes. Figure 1 shows the waveform and spectrogram representations of a typical token along with average spectra of the burst and vowel portions.

3. Results

English and Tamil data were analyzed separately. Linear mixed-effects models (using the `nlme` package in R) were fit to examine 1) the interaction effects of vowel du-

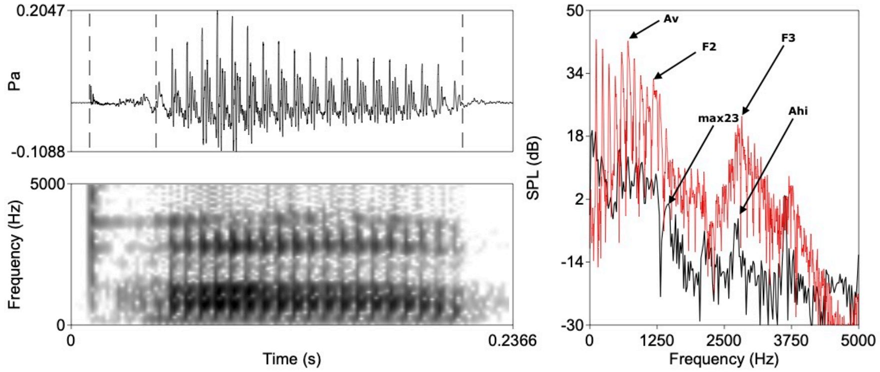

Figure 1: Waveform and spectrogram representations (left) of a typical "pa" token spoken by a male North American English speaker and averaged spectra for the burst and and vowel (right). Three dashed lines on the waveform represent burst onset, burst offset/vowel onset, and vowel offset. Critical spectral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

ration (proxy for speaking rate) and place of articulation on either F23Diff or HiDiff; 2) the relationship between voice-onset time, vowel duration, and place of articulation. All models were fit allowing for random slopes and intercepts for vowel duration by subject.

3.1. North American English: Relative burst amplitude

Across all speaking rates, bilabials had the highest F23Diff (*M*=32.53dB, *SD*=9.09 dB), followed by alveolars (*M*=26.31 dB, *SD*=6.79 dB) and velars (*M*=19.07 dB, *SD*=7.6 dB). Bilabials had the highest HiDiff (*M*=38.75 dB, *SD*=10.02 dB), followed by velars (*M*=31.18 dB, *SD*=7.24 dB) and alveolars (*M*=24.27 dB, *SD*=6.89 dB). Results of the model for the effect of vowel duration (as a function of place of articulation) on F23Diff are given in Table 1A. The model shows significant (p<0.001) fixed effects of place of articulation (*F*=803.16), vowel duration (*F*=135.47) and their interaction (*F*=54.01). Vowel duration had the most dramatic effect on F23Diff in bilabial stops, followed by velars, then alveolar stops. The model suggests that in each place of articulation, vowel duration has a negative effect on F23Diff. Model estimates are plotted, along with observed values in Figure 2A.

Table 1: Mixed effects models of relative amplitude of North American English stop bursts in a) the F2/F3 region (F23Diff) and b) above 3-3.5kHz (HiDiff), as a function of vowel duration and consonant place of articulation.

	A: F23Diff			B: HiDiff		
Predictor	β	*SE*	*t*	β	*SE*	*t*
POA-p	40.02	1.51	26.52**	46.92	1.36	34.39**
POA-t	28.83	1.52	18.93**	24.03	1.38	17.41 **
POA-k	23.29	1.53	15.21 **	34.40	1.38	24.76**
POA-p × V dur	-45.34	2.97	-15.24**	-52.26	5.24	-9.97**
POA-t × V dur	-13.70	3.09	-4.43**	-0.77	5.31	-0.14
POA-k× V dur	-22.98	3.10	-7.41 **	-18.19	5.32	-3.42*

Df (both models) = 2035. $^{**}p<0.0001$, $^{*}p<0.001$

The effects of vowel duration and place of articulation on HiDiff are given in Table 1B. All fixed effects were significant ($p<0.001$) as were the vowel duration interactions with bilabial and velar places of articulation, which suggests that as speaking rate increases the amplitude of the burst relative the vowel decreases. Vowel duration did not significantly affect the HiDiff of alveolar stops. Figure 2B plots model estimates of HiDiff along with actual data.

The overall effect of place of articulation on burst amplitude measures are consistent with those given in Stevens et al. (1999) and correspond to peak P_o given in Subtelny et al. (1966), with bilabials having the lowest amplitude (relative to the vowel) in both regions of the spectrum. Velars have a low-amplitude high-frequency prominence and high-amplitude mid-frequency range prominence. Alveolars have a mid-range amplitude higher, but similar to, bilabials, and a high-amplitude, high-frequency prominence. Curiously, the high-frequency burst amplitude of alveolars showed no significant effect of speaking rate, and the mid-frequency effect was significant but with a shallower slope (-13.70) than either bilabials (-45.34) or velars (-22.98).

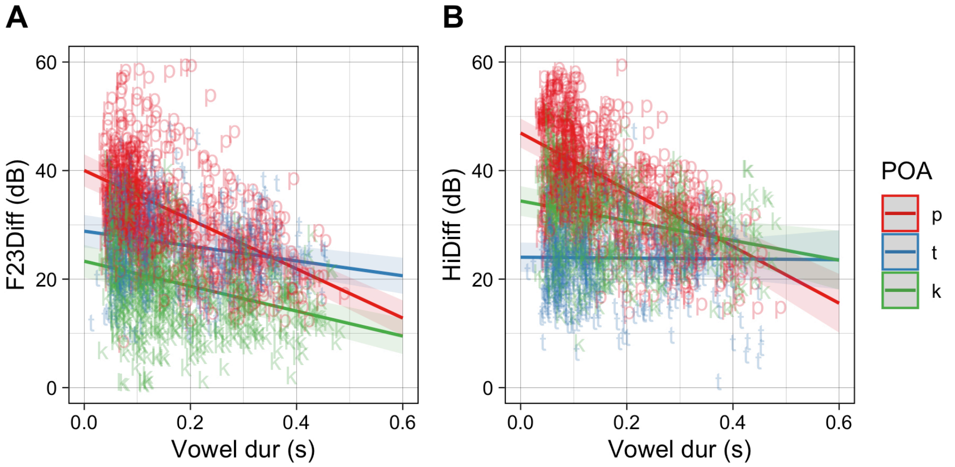

Figure 2: Predicted estimates (with confidence intervals) from linear mixed-effects models and observed values of A) F23Diff and B) HiDiff in the voiceless stops of North American English speakers.

3.2. Indian Tamil: Relative burst amplitude

Across all speaking rates, bilabials had the highest F23Diff (*M*=39.88 dB, *SD*=18.1 dB), followed by dentals (*M*=33.42 dB, *SD*=14.63 dB), retroflexes (*M*=26.52 dB, *SD*=9.25 dB), and velars (*M*=21.24 dB, *SD*=10.76 dB). Bilabials had the highest HiDiff (*M*=51.5 dB, *SD*=12.75 dB), followed by dentals (*M*=40.01 dB, *SD*=9.88 dB), velars (*M*=37.56 dB, *SD*=9.98 dB), and retroflexes (*M*=34.01 dB, *SD*=13.01 dB). The results of the model of F23Diff for the four places of articulation in Tamil is given in Table 2A. Both place of articulation (*F*=352.45) and vowel duration (*F*=7.41) were significant in the model ($p<0.001$) as was their interaction (*F*=2.69, $p<0.05$). As in the model of F23Diff in English, there was a negative effect of vowel duration in all places of articulation, reaching significance in all but velars. Bilabials showed the highest F23Diff, followed by dentals, then retroflexes, with velars having the lowest values. Model estimates are plotted, along with observed values in Figure 3A.

At frequencies above 3-3.5kHz (Table 2B), place of articulation had an effect on relative burst amplitude (*F*=297.32), while vowel duration alone did not have a significant effect (*F*=2.08). The interaction, however, between place of articulation and vowel duration was significant (*F*=3.01, $p<0.05$). Vowel duration negatively affected

Table 2: Mixed effects models of relative amplitude of Indian Tamil stop bursts in A) the F2/F3 region (F23Diff) and B) above 3-3.5kHz (HiDiff), as a function of vowel duration and consonant place of articulation.

	A: F23Diff			B: HiDiff		
Predictor	β	*SE*	*t*	β	*SE*	*t*
POA-p	44.98	4.27	10.52**	54.04	3.09	17.45**
POA-t̪	39.91	4.30	9.27**	40.62	3.13	12.94**
POA-k	23.98	4.31	5.56**	40.58	3.15	12.87**
POA-ʈ	30.33	4.32	7.01**	39.58	3.16	12.48**
POA-p × V dur	-27.98	10.18	-2.75**	-12.07	10.80	-1.12
POA-t̪ × V dur	-35.79	10.36	-3.45**	-4.53	10.98	-0.41
POA-k× V dur	-16.17	10.43	-1.55	-15.48	11.05	-1.40
POA-ʈ× V dur	-21.43	10.58	-2.02*	-27.46	11.02	-2.45*

Df (both models) = 1516. **$p<0.01$, *$p<0.05$

high-frequency burst amplitude in all places of articulation, reaching significance in only retroflexes. Table 2B gives the results of the linear model, showing the significant interaction effect only in retroflexes. Estimates and observed values are plotted in Figure 3B.

The overall effect of place of articulation on burst amplitude differs in the two languages. In general Tamil plosives have lower burst amplitudes (i.e., higher F23Diff and HiDiff) than English plosives, suggesting lower P_{o}.

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

To better understand the language differences in relative burst amplitude as a function of speaking rate, the relationship between voice onset time (VOT) and vowel duration was modeled in both North American English and Indian Tamil. The results of the VOT models for English and Tamil are shown in Figures 4A & B. For English, all fixed effects and interactions were significant ($p<0.001$), suggesting a clear postive relationship between vowel duration and VOT. This result was consistent with previous

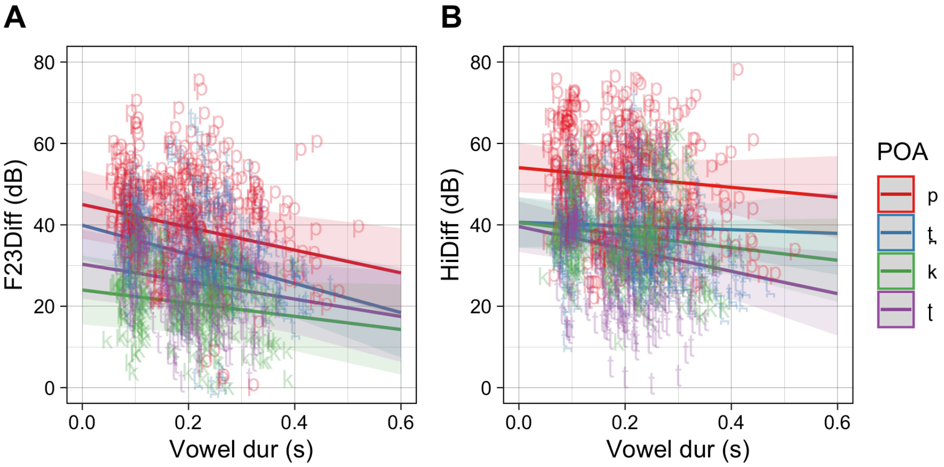

Figure 3: Predicted estimates (with confidence intervals) from linear mixed-effects models and observed values of A) F23Diff and B) HiDiff in the voiceless stops of Indian Tamil speakers.

research suggesting that as speaking rate decreases, vowel duration and VOT increase in equal proportions (Kessinger & Blumstein, 1997; Theodore et al., 2009).

For Tamil speakers, there was a main effect of place of articulation (F=310.88), vowel duration (F=6.45, p<0.05) and their interaction (F=21.40). The main interaction effect was driven primarily by velars (β=0.06, SE=0.014, df=1515, t=4.68, p<0.0001). Although there was a small, positive association between vowel duration and VOT in the other places of articulation, none achieved statistical significance (t=0-1). These results are largely consistent with literature suggesting a very small (and not statistically significant) effect of speaking rate on bilabial and alveolar VOT in short-lag languages (Thai and French) (Kessinger & Blumstein, 1997). That the VOT of velars in Tamil is significantly affected by vowel duration may be a consequence of the general tendency for velar plosives to have longer VOT values than other places of articulation due aerodynamic and biophysical reasons (Cho & Ladefoged, 1999; Kuehn & Moll, 1976).

Although there is an overall positive association between speaking rate and VOT in both languages, the size of the effect is reduced in the Tamil model owing to the phonologically specified short-lag implementation of VOT (75% of VOTs are under 0.025s),

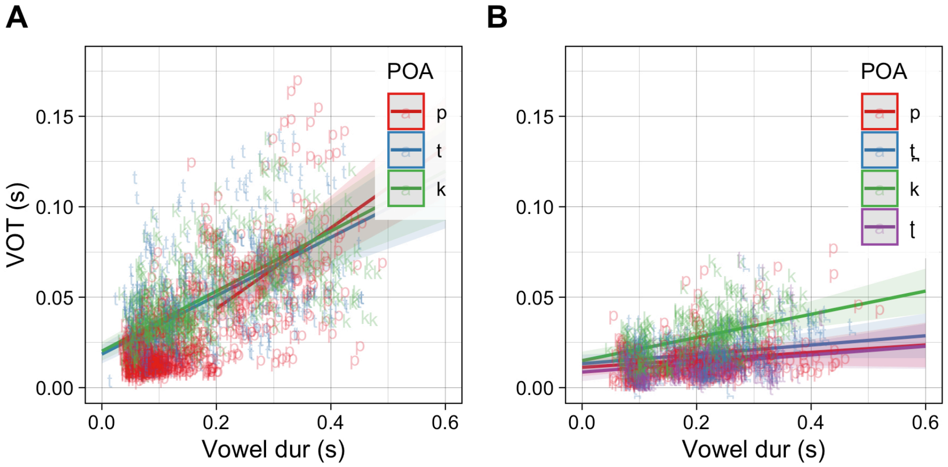

Figure 4: Predicted estimates (with confidence intervals) from linear mixed-effects models and observed values of VOT in A) North American English and B) Indian Tamil speakers.

while the long-lag VOT status in English allows for a more flexible oro-laryngeal timing. Given the relative immutability of VOT in Tamil, its effect on the two burst amplitude measures was modeled, independent of vowel duration in the two languages.

3.3.1. VOT effects on burst amplitude

Mixed effects models were fit to both English and Tamil HiDiff and F23Diff as a function of POA and VOT. All models allowed for random intercepts and slopes for VOT by subject. As expected, the English model of F23Diff showed significant effects of POA (F=535.15), VOT (F=55.33) and their interaction (F=35.76). Significant negative correlations were found for VOT in all three places of articulation (bilabial: β= -167.35, SE=18.06, t=-9.26; alveolar: β= -117.92, SE=18.75, t= -6.28; velar: β= -55.24, SE=19.69, t= -2.80). The model of HiDiff similarly showed significant effects of POA (F=709.05), VOT (F=19.70, and their interaction (F= 99.46). The effect was found in bilabials (β=-206.69, SE=29.79, t= -6.93) and velars (β=-72.02, SE=30.83, t= -2.33). There was no significant effect in alveolars. The results for English VOT effects on burst amplitude unsurprisingly mirror the speaking rate effects as VOT and speaking rate are highly correlated. This correlation is language-specific, as Tamil speakers did not necessarily increase VOT with increasing vowel duration.

Tamil models of F23Diff and HiDiff were likewise fit with POA and VOT as crossed predictors and random intercepts and slopes for VOT by subject. Unlike the English models, where effects of VOT and vowel duration were comparable, the Tamil models showed results from the vowel duration effects. For F23Diff, the main effect of POA was significant ($F(4, 1516)$=233.46) as was its interaction with VOT (F=20.53). The main effect of VOT was not significant. The model showed that VOT significantly affected F23Diff only in bilabials (β= -319.34, SE=81.24, t= -3.93). Similarly, the model of HiDiff (with the same predictors and random effects as the F23Diff model) had main effects of POA (F=303.53 and its interaction with VOT (F=20.53). The main effect of VOT was not significant. As with the F23Diff model, only the HiDiff of bilabials was significantly affected by VOT (β= -319.34, SE=81.24, t= -3.93).

4. General discussion

Linguistic factors such as syllable structure (Mackay, 1974) and phrasal length (Yuan et al., 2006), and extra-linguistic factors like emotional content of the speech (Mozziconacci & Hermes, 2000), and age and geographical background (Quené, 2005) of the speaker all have an effect on speaking rate. This study investigated whether speaking rate imposes constraints on the articulation of consonants via changes in burst amplitude, a proxy for intraoral pressure and glottal frication. Speaking rate affected burst amplitude in the current study, such that increasing rate (decreasing vowel duration) led to a decrease in amplitude (relative to vowel amplitude) in speakers' consonant bursts. While the effect is found in both North American English and Indian Tamil, it is markedly pronounced in English. This suggests that the differing implementation of oro-laryngeal timing in the two languages may be responsible for the variation. In this study, long-lag VOT in English voiceless stops is correlated with vowel duration as well as higher overall burst amplitudes than in Tamil. When coupled with the general tendency for voiceless aspirated stops to have long closure durations, it can be deduced that the high P_o associated with high burst amplitudes results from the phonological requirement of long-lag VOT in English. Conversely, Tamil initial stops, being short-lag, have lower burst amplitudes (relative to the vowel), suggesting lower P_o, with less pronounced effects of speaking rate. The study also revealed that the shared places of articulation in the two languages (bilabial and velar) have differing effects on relative burst amplitude in the two languages. The mid-frequency burst amplitude of bilabials in Tamil is lower than bilabials in English, while comparable differences are not found in velars.

4.1. Implications for sound change

Taken together, these findings may have particular import for understanding the role of burst amplitude in the perception of place of articulation. Although burst characteristics are part of a larger constellation of acoustic cues utilized by the listener in determining consonant identify, disruption (e.g., masking noise) of burst cues results in a greater reliance on formant transitional cues in perception (Dorman et al., 1977). When coupled with the tendency for speaking rate induced adjustments to formant transition patterns (Gay, 1978; Krull, 1989; Duez, 1992) as well as the association between fast speech and consonant lenition (e.g., Katz & Pitzanti, 2019; Priva & Gleason, 2020) we offer the possibility that reduced burst amplitude, triggered by increased rate, may be misperceived by the listener.

Further, speaking rate effects on burst amplitude are most obvious on bilabial plosives, suggesting that bilabials, which naturally have low amplitude bursts (Stevens et al., 1999) may be more prone to misperception. In languages with short-lag VOT, like Tamil, burst amplitudes generally, but especially in biabials, are naturally low. Indeed, bilabials, in consonant identification tasks in noise, have been shown to be disproportionately misidenitified as /h/ (Wang & Bilger, 1973; Woods et al., 2010). Further empirical investigation into the perceptual effect of both rate-induced burst amplitude reduction and low amplitude bursts in short-lag VOT languages may inform diachronic explanations of debuccalization of word-initial bilabials (p $\rightarrow$ h), especially in languages like Japanese (Sasaki, 2008) and Old Kannada (Gai, 1946) where there is no evidence of an intermediary labio-dental fricative stage. In this way, certain ‘seeds of sound change’ may be found in naturally conditioned variation in burst amplitude, owing to both phonological and aerodynamic forces.

References

Adams, S. G., Weismer, G., & Kent, R. D. (1993). Speaking rate and speech movement velocity profiles. *Journal of Speech, Language, and Hearing Research*, *36*, 41–54.

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

Bell-Berti, F. (1975). Control of pharyngeal cavity size for english voiced and voiceless stops. *The Journal of the Acoustical Society of America*, *57*, 456–461.

Blumstein, S. E., & Stevens, K. N. (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*, 1001–1017.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*, 207–229.

Chodroff, E., & Wilson, C. (2014). Burst spectrum as a cue for the stop voicing contrast in American English. *The Journal of the Acoustical Society of America*, *136*, 2762–2772.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, *22*, 109–122.

Dromey, C., & Ramig, L. O. (1998). The effect of lung volume on selected phonatory and articulatory variables. *Journal of Speech, Language, and Hearing Research*, *41*, 491–502.

Duez, D. (1992). Second formant locus-nucleus patterns: an investigation of spontaneouos french speech. *Speech Communication*, *11*, 417–427.

Gai, G. S. (1946). *Historical Grammar of Old Kannada: Based Entirely on the Kannada Inscriptions of the 8th, 9th and 10th Centuries AD* volume 1. Deccan College, Postgraduate and Research Institute.

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

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica*, *38*, 148–158.

Haag, W. K. (1977). Oral Air Pressure and Airflow in German Stop Consonants. *Phonetica*, *34*, 31–39.

Hertegård, S., Gauffin, J., & Lindestad, P.-Å. (1995). A comparison of subglottal and intraoral pressure measurements during phonation. *Journal of Voice*, *9*, 149–155.

Katz, J., & Pitzanti, G. (2019). The phonetics and phonology of lenition: A campidanese sardinian case study. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *10*.

Keane, E. (2004). Tamil. *Journal of the International Phonetic Association*, *34*, 111–116. doi:`10.1017/S0025100304001549`.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in thai, french, and english. *Journal of Phonetics*, *25*, 143–168.

Krause, J. C., & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *The Journal of the Acoustical Society of America*, *112*, 2165–2172.

Krull, D. (1989). Second formant locus patterns and consonant-vowel coarticulation in spontaneous speech. *Phonetic Experimental Research at the Institute of Linguistics, University of Stockholm*, *10*, 87–108.

Kuehn, D. P., & Moll, K. L. (1976). A cineradiographic study of vc and cv articulatory velocities. *Journal of Phonetics*, *4*, 303–320.

Laukkanen, A.-M., Vilkman, E., Alku, P., & Oksanen, H. (1996). Physical variations related to stress and emotional state: a preliminary study. *Journal of Phonetics*, *24*, 313–335.

Lindblom, B. (1983). Economy of speech gestures. In P. MacNeilage (Ed.), *The Production of Speech* (pp. 217–245). Springer.

Mackay, D. G. (1974). Aspects of the syntax of behavior: Syllable structure and speech rate. *Quarterly Journal of Experimental Psychology*, *26*, 642–657.

Miller, C. J., & Daniloff, R. (1977). Aerodynamics of stops in continuous speech. *Journal of Phonetics*, *5*, 351–360.

Mozziconacci, S. J., & Hermes, D. J. (2000). Expression of emotion and attitude through temporal speech variations. In *Sixth International Conference on Spoken Language Processing*.

Müller, E. M., & Brown Jr, W. S. (1980). Variations in the supraglottal air pressure waveform and their articulatory interpretation. In *Speech and Language* (pp. 317–389). Elsevier volume 4.

Murry, T., & Brown Jr, W. (1976). Peak intraoral air pressures in whispered stop consonants. *Journal of Phonetics*, *4*, 183–187.

Ohde, R. N., & Stevens, K. N. (1983). Effect of burst amplitude on the perception of stop consonant place of articulation. *The Journal of the Acoustical Society of America*, *74*, 706–714.

Picheny, M., Durlach, N., & Braida, L. (1986). Speaking clearly for the hard of hearing. ii: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29 4*, 434–46.

Pickett, E. R., Blumstein, S. E., & Burton, M. W. (1999). Effects of speaking rate on the singleton/geminate consonant contrast in italian. *Phonetica*, *56*, 135–157.

Priva, U. C., & Gleason, E. (2020). The causal structure of lenition: A case for the causal precedence of durational shortening. *Language*, *96*, 413–448.

Quené, H. (2005). Modeling of between-speaker and within-speaker variation in spontaneous speech tempo. In *Ninth European Conference on Speech Communication and Technology*.

Rothenberg, M. (1966). *The Breath-Stream Dynamics of Simple-Released-Plosive Production*. Ph.D. thesis University of Michigan.

Sasaki, K. (2008). Hardening alternation in the Mitsukaido dialect of Japanese. *Gengo Kenkyu*, *134*, 85–118.

Stevens, K. N., Manuel, S. Y., & Matthies, M. (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, J. D., Worth, J. H., & Sakuda, M. (1966). Intraoral pressure and rate of flow during speech. *Journal of Speech and Hearing Research*, *9*, 498–518.

Sundberg, J. (2018). The singing voice. *The Oxford handbook of voice perception*, (p. 117).

Theodore, R. M., Miller, J. L., & DeSteno, D. (2009). Individual talker differences in voice-onset-time: Contextual influences. *The Journal of the Acoustical Society of America*, *125*, 3974–3982.

Wang, M. D., & Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *The Journal of the Acoustical Society of America*, *54*, 1248–1266.

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *The Journal of the Acoustical Society of America*, *73*, 1322–1336.

Woods, D. L., Yund, E. W., Herron, T. J., & Cruadhlaoich, M. A. U. (2010). Consonant identification in consonant-vowel-consonant syllables in speech-spectrum noise. *The Journal of the Acoustical Society of America*, *127*, 1609–1623.

Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. In *Ninth International Conference on Spoken Language Processing*.