Rate Effects on Swedish VOT: Evidence for Phonological Overspecification

(author names are alphabetical)

Jill Beckman[1]
Pétur Helgason[2]
Bob McMurray[1*]
and
Catherine Ringen[1]

[1]University of Iowa, Iowa City, IA 52242, USA
[2]Uppsala University, SE-751 05, Uppsala, Sweden

Corresponding Author:
Bob McMurray
E11 SSH
Dept. of Psychology
University of Iowa
Iowa City, IA 52242
USA
bob-mcmurray@uiowa.edu
Tel: 319-335-2408

Manuscript Accepted for Publication, *Journal of Phonetics*

Abstract

Previous research has found asymmetric effects of speaking rate on VOT cross-linguistically: as rate slows, long-lag VOTs and negative VOTs increase, but short-lag VOTs remain essentially unchanged. If we assume, as have many phonologists, that the two-way contrast in voicing languages (e.g. French) is [voice] vs. [∅] and in aspirating languages (e.g. English) is [spread glottis] vs. [∅], then it appears that at slower rates, a phonological contrast is heightened by selective increase in the phonetic cue for the specified feature. Thus, slowing down causes longer aspiration in aspirating languages and longer prevoicing in voicing languages but no change in short-lag stops. We report the results of an experiment on Central Standard Swedish stops designed to investigate the effect of speaking rate on VOT. CS Swedish uses both prevoiced and aspirated stops in utterance-initial position, hence the phonological feature(s) involved in this contrast is not clear. We found that both prevoicing and aspiration increase in slow speech in Swedish. This suggests that *both* [voice] *and* [spread glottis] are the specified features of phonological contrast in CS Swedish, and in turn raises questions about whether phonological specification more generally is economical. Moreover, the fact that speaking rate affects VOT even in situations like CS Swedish in which the phonological contrast is over-specified suggest that such modification is largely due to production dynamics, not speakers' sensitivity to listeners' needs.

Keywords

Voicing, Voice Onset Time, Swedish, Phonological Features,
Privative Features, Underspecification Theory

Rate Effects on Swedish VOT: Evidence for Phonological Overspecification

1.0 Introduction

Many languages exhibit contrasts that are informally described as voicing contrasts, though there is considerable debate about the underlying phonetic and phonological make up of these contrasts. English, for example, has a two-way contrast between aspirated (fortis) stops, which are informally described as "p, t, k", but phonetically appear as [p^h], [t^h], and [k^h], and lenis stops, which are informally described as "b, d, g", but which appear phonetically as the voiceless unaspirated stops [p, t, k]. French also has a two-way contrast, though it contrasts prevoiced (lenis) stops (in which voicing precedes the release of the closure) and plain unaspirated (fortis) stops. Further, languages like Thai and Eastern Armenian have three-way contrasts, distinguishing both prevoiced and aspirated stops from voiceless unaspirated stops (Table 1).

These differences in stop contrasts can be quantified phonetically as differences in Voice Onset Time, or VOT (Lisker & Abramson, 1964), which refers to the time difference between the release of the closure and the onset of vocal fold vibration. VOT for utterance-initial, prevocalic stops is easy to measure from acoustic recordings. In terms of VOT, prevoiced stops have negative VOTs in the −40 to −80 ms range, with voicing appearing before the release of the stops; short-lag stops have VOTs near 0; and aspirated stops are marked by between 40 and 80 ms of aspiration after the release of closure.

Until recently, it was thought that languages employ the minimal number of distinctive features necessary to encode the number of contrasts. That is, a language with a two-way contrast in voice (henceforth a voicing language) employs the feature [voice], which commonly manifests as prevoicing (as in French) and a language with a two-way contrast in aspiration (henceforth an aspirating language) employs the feature [spread glottis] ([sg]) (as in English). However, no language with a two-way contrast employs both (Iverson & Salmons, 1995).

In contrast, recently, Helgason and Ringen demonstrated that Central Standard (CS) Swedish has both prevoiced and aspirated stops (Helgason & Ringen, 2008), like Thai. But unlike Thai, it has *only* a two-way contrast. Thus, the contrast in CS Swedish is over-specified, potentially using both [voice] and [sg] to mark a two-way distinction. In terms of VOT, while the typical VOT difference between fortis and lenis stops in both aspirating and voicing languages might be approximately 60 ms, in CS Swedish, it is approximately 120 ms. This raises the question of what phonological features specify the contrast in CS Swedish, and makes an investigation of the phonetic properties of such a system particularly important.

In the present paper, we examine speaking rate. As we will describe, phonetic data from a number studies suggest that speaking rate affects VOT, but typically in only one of the two categories of stops (Miller, Green & Reeves, 1986; Pind, 1995; Kessinger & Blumstein, 1997; Allen & Miller, 1999; though see Magloire & Green, 1999). Further, across studies

Table 1: Summary of the stop categories used in several languages. Shown is the phonetic symbol and typical VOT range of each category.

# Categories	*Language*	*Prevoiced*	*Short-lag*	*Aspirated*
2	English Icelandic	---	p 0 to 10 ms	p^h +40 to +80 ms
2	French Spanish	b −80 to −40 ms	p 0 to 10 ms	---
3	Thai Eastern Armenian	b −80 to −40 ms	p 0 to 10 ms	p^h +40 to +80 ms
2	CS Swedish	b −40 to −80 ms	---	p^h +40 to +80 ms

we see that which category is affected appears to mirror the features used by the phonology of a language, making this an important laboratory manipulation for understanding which voicing features may be involved in CS Swedish. Most pertinently, Kessinger and Blumstein (1997) found the effect of speaking rate on VOT production was not the same for all categories. For French and Thai (both of which use the [voice] feature), as speaking rate decreased, the amount of prevoicing increased. By contrast, for Thai and English (which use the [sg] feature), the amount of aspiration increased with slower speaking rates. Importantly, in all three languages, there was little or no change in the short-lag stops (which are [−voice] [−sg] in all three).

This suggests a strong parallel with the phonology of the language: in slower speech, the duration of the phonetic cue increases, but only for the specified or marked feature. That is, slowing down yields longer aspiration in English and Thai because [sg] is a specified feature in both English and Thai; slowing down results in longer prevoicing in French and Thai because [voice] is a specified feature in both French and Thai.

The present study builds on these results to examine how speaking rate affects VOT in CS Swedish. As we shall describe, this has important implications for understanding the phonology of the laryngeal contrast in CS Swedish, and this in turn speaks to whether languages can adopt a non-minimal feature-set. At a phonetic level, CS Swedish also offers an intriguing platform on which to understand how speakers highlight phonetic contrasts via manipulations of the speech mode by asking participants to slow down or produce clear speech. That is, speakers may increase the distance between the categories (in terms of VOT) despite the fact there is apparently no need in CS Swedish, because even in fast speech there is more contrast than is necessary.

In Section 2, we present a brief summary of prior work on the effect of speaking rate on VOT production and perception, along with Kessinger & Blumstein's (1997) results. We also consider how these results can be understood in terms of laryngeal features. We next present our phonetic experiment and the results. Finally, we discuss the implications of these results for phonological features.

2. Background.

VOT is fundamentally a temporal variable, a time difference between two articulatory events. As a result, there has been considerable interest in the relationship of this variable to speaking rate. Summerfield (1981) was the first to investigate this, finding that English listeners shifted their perceptual boundary on the VOT continuum as a function of speaking rate (cued either by a preceding sentence or the length of the vowel in the CV). At fast rates, listeners shifted their category boundary such that intermediate VOTs tended to be classified as "p", while at slow rates of speech, intermediate VOTs were classified as "b" (See also Miller & Volaitis, 1989; Volaitis & Miller, 1992; Pind, 1995; McMurray, Clayards, Tanenhaus & Aslin, 2008).

Motivated in part by these perceptual results, subsequent work measured the effect of speaking rate on the VOTs that speakers produced (Miller, Green & Reeves, 1986; Pind, 1995; Kessinger & Blumstein, 1997; Allen & Miller, 1999; Magloire & Green, 1999). The initial work on English suggested that VOTs in the two types of stops change asymmetrically with changes in speaking rate. *As speech slowed, VOT increased in aspirated (voiceless) stops, but VOT in the short-lag stops was largely unchanged*. Moreover, this is not an isolated property of English; it also occurs in Icelandic (Pind, 1995), another aspirating language.

Kessinger and Blumstein (1997) present the most comprehensive cross-linguistic analysis. They tested the effects of speaking rate on the production of VOT in English, French, and Thai. Speakers produced minimal or near-minimal pairs of words in three conditions: isolation (word list), at a slow rate (carrier phrase, demonstrated as "clearly enunciated speech") and at a fast rate (carrier phrase, produced "as quickly as possible without forsaking accuracy"). Importantly, Kessinger and Blumstein also found asymmetric effects of speaking rate on *VOTs in all three languages.* In English, as in previous studies, long-lag VOTs increased at slower rates, while short-lag VOTs remained unchanged; in French, the negative VOTs increased at slower rates, while again, the short-lag category remained unchanged; and in Thai, both the negative and the long-lag VOTs increased in duration, while the short-lag category remained unchanged.

Similar findings were also observed by Magloire and Green (1999). They studied speakers of Spanish (a language that uses prevoicing), speakers of English, and bilinguals in a similar paradigm. For English monolinguals, the short-lag stops exhibited little change as a result of speaking rate, while the long-lag stops were affected. Spanish speakers showed a large increase in prevoicing, and a smaller change for short-lag stops (a difference of 10 ms of VOT for short-lag stops vs. 20 ms for prevoiced stops). Bilinguals showed the pattern of the language they were speaking at that time.

Thus, there appears to be fairly robust evidence that the VOT of prevoiced and aspirated stops is affected by speaking rate, but that short-lag stops are either not affected at all, or are only minimally affected. Moreover, such effects are not limited to speaking rate manipulations: studies examining VOTs in clear speech or as a function of prosodic stress suggest strong parallels. This is relevant because words spoken in clear speech mode or in prosodically strong positions undergo acoustic modifications similar to those of words spoken at a slower rate (see Bradlow & Smiljanic, 2009, for a review), though these are clearly independent manipulations (Krause & Braida, 2002). For example, Smiljanic and Bradlow (2008) found that in the clear speech mode, Croatian speakers lengthen prevoiced stops, leaving short-lag stops unchanged; while English listeners lengthen long-lag VOTs, again leaving short-lag stops unchanged. Thus, whatever mechanism is responsible for these asymmetric changes may extend beyond speaking rate and shed light on how speakers' intention to highlight a phonetic contrast affects segmental properties.

The asymmetry in effects on short- and long-lag (or prevoiced) VOTs may seem to require no explanation*:* after all, what could a VOT near 0 ms change into? However, a number of possibilities are available. Such VOTs could move away from the other category. Some English speakers use prevoicing (Keating, 1984), and this could certainly be lengthened; likewise, Spanish or French speakers (particularly bilinguals who are used to longer periods of aspiration) could lengthen the aspiration of these stops (e.g. Magloire & Green). Alternatively, in English, all VOTs could be lengthened, a form of syntagmatic enhancement (increasing the contrast between the consonant and vowel) discussed in work on prosodic effects on articulation (e.g., Cho & Jun, 2000) and infant-directed speech, another type of clear speech mode (Englund, 2005). Thus, it is by no means obvious that short-lag VOTs would remain unchanged with changes in speaking rate across languages that use this category of stops differently.

2.1 Explanations for the asymmetry.

There have been few concrete theoretical proposals to account for the asymmetrical effects of speaking rate on VOT. Kessinger and Blumstein (1997) suggest that speakers may be actively holding one category constant in order to preserve the phonetic contrast. That is, in aspirating languages, if the short-lag VOT lengthened at slow rates, it might be confused with a short aspirated VOT. By holding one category constant, listeners would have an anchor. This is similar to arguments made in the large literature on clear-speech and stress effects on vowel length. Similar to VOT, phonemically short vowels tend to lengthen less than long vowels in stressed syllables or clear speech (e.g., de Jong and Zawaydeh, 2002; Heldner & Strangert, 2001; Pind, 1999; see Smiljanic & Bradlow, 2008, for a review). As with VOT, this has been hypothesized to help preserve the contrast by allowing listeners to maintain a more fixed boundary between short and long vowels (while simultaneously moving the long vowels away from it for better discriminability).

However, as Kessinger and Blumstein (1997) point out, this explanation is not sufficient to account for VOT. Their own results on prevoicing in French (Kessinger & Blumstein, 1997) and the later results on Spanish (Magloire & Green, 1999) suggest that the short-lag VOTs in these languages could be safely lengthened, without requiring listeners to alter a phonetic boundary, and yet speakers do not lengthen them[1]. Kessinger and Blumstein also suggest an articulatory account: lengthening a short-lag VOT requires an additional gesture (aspiration). In aspirating languages, of course, this would reduce or eliminate the contrast, while in prevoicing languages this additional gesture would simply be additional effort. Either way, this would result in a preference toward not adjusting the short-lag stops.

If we consider the phonological implications of this articulatory account, then it may explain the close correspondence between which categories show rate effects and the phonology of these languages. That is, let us assume that speakers are actively attempting to achieve *voicing* in the voiced stops in French and Thai and actively attempting to achieve *aspiration* in the aspirated stops in English and Thai. One of the reasons that speakers slow down, of course, is to make their speech easier to understand. Thus when the rate is slower, then speakers are able to produce more of whatever acoustic property they are trying to produce. So, if the features specified in the phonology reflect speakers' goals, then we might expect a speaking rate to affect the phonetic cues associated with the features [voice] and [sg] (aspiration), but to not affect categories defined by their absence.

An alternative account is that speakers are just attempting to increase the contrast. If what speakers were attempting to do were to increase the contrast, then it is difficult to see why the changes in English and French are not symmetrical. Speakers could increase the contrast in English by prevoicing the short-lag stops and aspirating the long-lag stops, or speakers in French could increase the contrast by increasing the positive VOT in the short-lag stops and increasing the prevoicing in the others. Instead, their behavior is more selective – even bilinguals do just one or the other, depending on the language being spoken (Magloire & Green, 1999). Thus, accounts emphasizing the contrast between features, rather than the featural identity of the contrast, cannot explain these results. Indeed, this is supported by perceptual work showing much larger changes in the center and extreme edges of speech categories than at the boundary (Miller & Volaitis, 1989; Miller, O'Rourke & Volaitis, 1997) – listeners are apparently adjusting their prototypes of the categories more than the boundaries.

Our account, based on the intended features, is in line with the idea that phonologically, laryngeal features are *privative*—that is*,* that is, they are defined by the presence or absence of a gesture—rather than binary—defined by two values with equal status. Privativity has gained support among phonologists for a purely phonological reason: it is, apparently, never necessary to refer to the minus value of a feature in the phonology—that is, [−voice] is never "active" in phonology, suggesting that it isn't there (c.f. Mester & Itô 1989).

If features are privative, there is a clear relation between the phonological feature and the phonetic cue: [voice] will have the cue of prevoicing in initial position and [sg] will implicate aspiration. Thus, the two-way contrast in languages such as English is between stops specified as [sg] and [∅], and in languages such as French, the contrast is between stops that are specified as [voice] and [∅] (Anderson & Ewen, 1987; Beckman, Jessen & Ringen, 2009; Harris, 1994; Iverson & Salmons,

---

[1] *It could also be argued that increasing the VOT in a pre-voicing language may result in stops that cross the hypothesized psychophysical boundary at +20 ms of VOT. Such a discontinuity has been hypothesized on the basis of psychoacoustic work with humans (e.g., Pisoni, 1977), speech perception work with animals (Kuhl & Miller, 1975) and electrophysiological work with humans (e.g., Sharma & Dorman, 1999). While prevoicing languages don't have a second category there, if such a boundary is present, extending the VOT could still cause problems for listeners. However such work has been controversial in all three domains (e.g., Soli, 1983; Ohlemiller, Jones, Heidbreder, Clark & Miller, 1999) and recent electrophysiological work suggests that even English listeners (who should clearly show a discontinuity in VOT perception) do not show such a discontinuity at +20 ms (Frye, McGraw-Fisher, Coty, Zarella, Liederman, & Halgren, 2007; Toscano, McMurray, Dennhardt & Luck, 2010). This coupled with robust evidence of flexible category boundaries (e.g., Repp, 1982) and significant within-category sensitivity (Massaro & Cohen, 1983; McMurray, Tanenhaus & Aslin, 2002) suggests there may not be a psychoacoustic discontinuity for VOT, making this hypothesis untenable.*

1995; Jessen & Ringen, 2002; Honeybone, 2005; Petrova et al., 2006). The three-way stop contrast of Thai is defined by both features, [sg] and [voice], as well as [∅] (Iverson & Salmons, 1995). In all three languages, the short-lag stops are [∅].

On this view of laryngeal specification, the foregoing results on speaking rate make sense: in slower speech, the duration of the phonetic cue for the specified feature is increased. That is, slowing down yields longer aspiration in English, Icelandic and Thai because [sg] is a specified feature in all three languages. Similarly, slowing down results in longer prevoicing in Spanish, French and Thai because [voice] is a specified feature in these languages. The fact that the voiceless unaspirated stops in all three languages remain unchanged provides some support for the phonologists' claim that this is the unmarked category. Critically, this also offers a framework for thinking about the effects of hyperarticulation on segmental cues. While some have argued that hyperarticulation is intended to amplify phonetic contrasts for the benefit of the listener (e.g., Lindblom, 1990), this is difficult to reconcile with the systematic yet asymmetrical effects on VOT—why do speakers selectively amplify a single category? In contrast, if hyperarticulation simply enables speakers to produce more of the privative feature they are aiming for, such effects can be related to the phonology.

CS Swedish has a two-way stop contrast between initial stops that are prevoiced and aspirated (Helgason & Ringen, 2008; Ringen & Helgason, 2004). Thus, the phonological features involved in this contrast are not clear, as both fortis and lenis stops appear to have the phonetic properties of one of the marked categories. This leaves three possible analyses. First, perhaps voiced stops are phonologically specified, while voiceless stops are unmarked (e.g. [voice] vs. [∅]). Second, a complementary analysis is that voiceless stops are specified, and voiced are unmarked ([sg] vs. [∅]). Finally, it might be argued that both [sg] and [voice] are phonological features of CS Swedish stops, but that there are no stops specified as [∅] (Beckman & Ringen, 2004; Ringen & Helgason, 2004).

This, in turn, raises the question of speaking rate. If VOT cues associated with specified laryngeal features are lengthened in slower speech, then each analysis of CS Swedish offers three possible predictions. 1) If [voice] is the only marked feature in CS Swedish, slower speech should result in only the prevoicing, but not the aspiration, being lengthened. 2) If only [sg] is marked, then aspiration, but not prevoicing, will increase at slower speaking rates. 3) If speakers are actively trying to achieve both, then both prevoicing and aspiration increase in slow speech.

While these specific predictions largely concern the phonology of CS Swedish, there are also several broader issues to consider. First, if we find movement in either of the categories, particularly if this movement is of a similar size to that seen in other languages, this would suggest that speaking rate effects are somewhat unrelated to speakers' intentions to increase the differences between the categories—this contrast is already over-specified in CS Swedish, even in fast speech. Rather, such changes are better described as changes in a single category or feature, and may be independent of the "system" of laryngeal contrasts, or of listener's ability to distinguish them. Second, if we find differences in both types of stops, that is, if CS Swedish speakers are trying to achieve both [sg] and [voice], why?

To start to understand these issues, we conducted a measurement study to determine how speaking rate affects both voiced and voiceless stops in CS Swedish.

3.0 Method.

3.1 Subjects

Eight subjects (four males and four females) were paid for participating in this experiment. All were speakers of the Central Standard variety of Swedish. Subjects' ages ranged from early twenties to mid-fifties. All subjects had been born and raised in Stockholm or areas surrounding Stockholm. Apart from one subject who spent one year in England in her early twenties, none of the subjects have lived outside the area in which CSS is spoken. All reported having normal hearing.

Table 2: Stimuli used in the Experiment.

Fortis Stops			***Lenis Stops***		
Word	**IPA**	**Gloss**	**Word**	**IPA**	**Gloss**
buss	[bɵsː]	‘bus’	puss	[pʰɵsː]	‘kiss’
ball	[balː]	‘trendy, cool’	pall	[pʰalː]	‘stool, pallet’
boss	[bɔsː]	‘boss’	post	[pʰɔst]	‘post, mail’
dum	[dɵmː]	‘stupid’	tum	[tʰɵmː]	‘inch’
dagg	[dagː]	‘dew’	tagg	[tʰagː]	‘barb, thorn’
dolk	[dɔlk]	‘dagger’	tolk	[tʰɔlk]	‘interpreter’
guld	[gɵld]	‘gold’	kull	[kʰɵlː]	‘litter’
gall	[galː]	‘bile formation’	kall	[kʰalː]	‘cold’
golv	[gɔlv]	‘floor’	kolv	[kʰɔlv]	‘piston’

3.2 Stimuli

The target words contained fortis stops and lenis stops in word-initial position. Three stop places of articulation were considered: bilabial, dental and velar[2]. The stimuli consisted of a set of 9 minimal or near-minimal pairs, yielding a total of 18 target words (see Table 2). Each stop place was represented by three different word pairs; and each word pair used a different vowel ([ɵ], [a] and [ɔ]). All the target words had a CVC: or CVCC structure.

3.3 Procedure

The target words were elicited in three different conditions. First, the subjects read a list of isolated words in a word list without specific instructions as to speaking rate (“*Isolated”).* Second, the subjects read a series of carrier phrase sentences (*Läs* ____ *igen* ‘Read ____ again’) in which the target words were embedded, again without specific instructions as to speaking rate (“*Slow”).* Third, the subjects read the same series of carrier phrase sentences, but were asked to increase their speaking rate without sacrificing accuracy (“*Fast”*). For all three conditions, the target words were arranged in a pseudo-random order using fillers to disguise the nature of the experiment. Preliminary analyses did not reveal substantive differences between the *Isolated* and *Slow* condition, so the *Isolated* condition was dropped to permit a more straightforward comparison of speaking rates.

When participants are asked to read lists of words or sentences without instructions as to speaking rate, a fairly slow speaking rate typically results. In contrast, when the subjects are told to speed up, they come closer to a speaking rate that one finds in unscripted conversational speech. In Helgason’s (2002) unscripted, spontaneous speech data from CSS, the estimated speaking rate ranged between approximately 5 and 6 syllables per second for the four subjects examined. In the present data, the syllable duration in the “slow” condition averaged 436 ms (range: 267-658 ms), and the fast condition averaged 337.7 ms (range: 207-603 ms), which translates to 2.3 syllables per second in the slow condition, and 3.0 syllables per second in the fast condition. This is not as fast as the unscripted results from Helgason (2002), but this should be expected since the words occurred in a more prominent part of the sentence.

Data were recorded in a sound-treated room at the phonetics lab at Stockholm University, using a Marantz PMD670 solid-state recorder and Sennheiser MKE2 microphone. The microphone headset was

[2] *Retroflex stops do not occur utterance-initially in Swedish.*

mounted with placement approximately 3 cm out and to the side of the corner of the subject's mouth. The data were sampled at 44.1 kHz and subsequently down-sampled to 22.05 kHz for analysis.

VOT was measured for all word-initial stops. Negative VOT (prevoicing) was measured as the interval from voice onset to the release of the stop. The onset of periodic oscillation indicating vocal fold vibration was identified visually from oscillograms and spectrograms. Positive VOT (postaspiration) was measured as the interval from stop release to the onset of modal voicing.[3] We also measured word duration, as a proxy for overall speaking rate. Because all our test words were embedded in carrier phrase sentences and were acoustically easy to separate from their context, estimating word duration proved unproblematic.

4. Results

The primary question of interest was whether our speaking rate manipulation affected the VOTs of both fortis and lenis stops. If both stops are specified in CS Swedish, we might expect to see VOTs lengthen for both when the speaking rate is slowed. In contrast, if only one is specified, the other should be relatively constant with respect to speaking rate manipulations.

Before asking this, however, we first asked if the rate manipulation itself had the intended effect on speaking rate. Thus, we examined syllable length as a function of the speaking rate condition. Next we assessed our primary question—the effect of speaking rate on prevoiced and voiceless VOTs. Our design included all three places of articulation (labial, coronal, velar), and three vowels ([a], [ɔ], [ɵ]), thus we also examined the effect of each of these factors on VOT, primarily to establish that our speaking rate effects can be seen in all six contexts.

4.1 Word length (global rate)

The first analysis asked whether our rate manipulations resulted in different speaking rates. We used the total word duration in ms as a proxy for speaking rate in a repeated measures ANOVA examining rate condition (slow or fast) along with the secondary variables, place (labial, coronal, velar), vowel ([a], [ɔ], [ɵ]) and stop-type (fortis-lenis).

Most importantly, we found a main effect of speaking rate condition (F(1,7)=107.9, p<.0001; see Figure 1). In fast speech, words were spoken at an average of 337.9 ms (SD=36.6), while in slow speech they averaged 436.7 ms (SD=46), an almost 100 ms

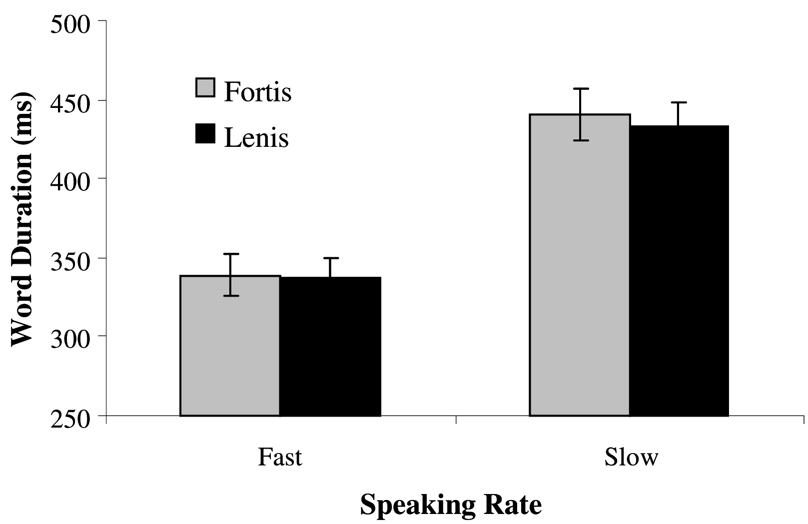

Figure 1: Effect of speaking rate and voicing on word duration.

[3] *By using onset of modal voicing rather than voice onset proper we are following the suggestion of Ladefoged & Maddieson (1996: 70). They define aspiration as "a period after the release of a stricture and before the start of regular voicing [...] in which the vocal folds are markedly further apart than they are in modally voiced sounds". Thus breathy voicing is regarded as a part of the aspiration. Therefore, our use of VOT actually refers to modal voice onset time rather than voice onset time proper. The onset of modal voicing was determined by visually identifying the cessation of aperiodicity in the mid-range of the spectrum, indicating the return to "regular" voicing for a given subject.*

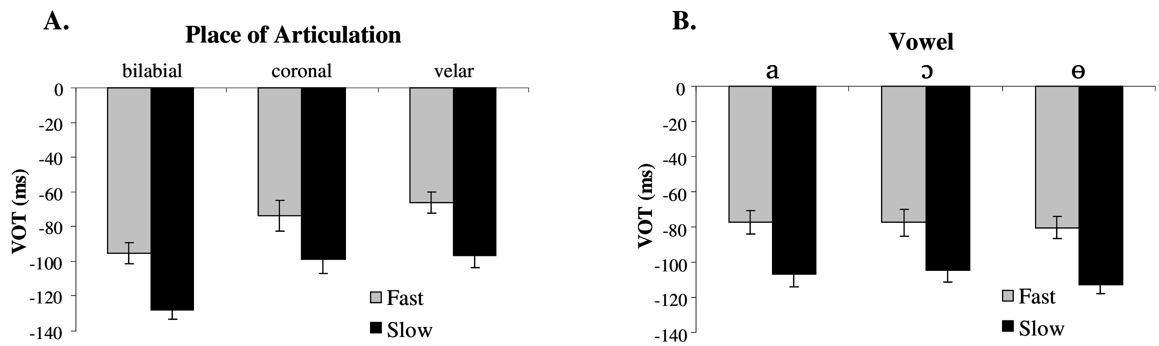

Figure 2: Effect of speaking rate on VOT for lenis stops A) broken down by place of articulation; B) broken down by vowel.

difference. Also important, speaking rate did not interact with voicing ($F(1,7)=1.6$, $p=0.24$). The non-significant interaction means that the difference in word duration as a function of speaking rate was similar in both the voiced and voiceless stops. The presence of such an interaction would have confounded a comparison of speaking rate effects on VOT across fortis and lenis stops. Finally, speaking rate showed no two-way interactions with any other variable (Rate x Place: $F<1$; Rate x Vowel: $F(2,14)=2.8$, $p=0.10$) and no three-way or four-way interactions (Rate x Voicing x Vowel: $F<1$; Rate x Place x Vowel: $F<1$; Rate x Voicing x Place: $F(2,14)=1.8$, $p=0.20$; Four-Way: $F<1$).

There were a number of other effects on word length that, while not central to the present question, were also examined. There was a marginal effect of voicing on word length ($F(1,7)=3.7$, $p=0.096$), with fortis stops (M=389.6 ms) slightly longer than lenis , (M=384.9).. Place of articulation was highly significant ($F(2,14)=17.3$, $p=0.0002$), as was vowel ($F(2,14)=82.5$, $p<.0001$). In general, words beginning than bilabials were longer than those beginning with coronals and velars, and [ɔ]'s were longer than [a] and [ɵ].

Voicing also showed a marginal interaction with place of articulation ($F(2,14)=3.6$, $p=0.055$), due to fact that fortis stops were slightly longer than lenis for bilabials ($\Delta=6.2$ ms; $T(7)=1.9$, $p=.10$) and coronals ($\Delta=13.8$; $T(7)=2.7$, $p=.031$), but this was reversed for velars ($\Delta=-6.2$; $T<1$). Voicing did not interact with vowel ($F(2,14)=1.6$, $p=0.23$). The place x vowel interaction was significant ($F(4,28)=43.6$, $p<.0001$) due to the fact that for bilabials, the effect of vowel ([ɔ] longer than [a], [ɵ]) was similar but smaller ($F(2,14)=5.5$, $p=.018$) than for coronals ($F(2,14)=221.0$, $p<.0001$) and velars ($F(2,14)=31.3$, $p<.0001$). Finally, the three-way interaction of voicing, place and vowel was significant ($F(4,28)=16.5$, $p<.0001$). All of these effects of place, vowel and their interactions are likely due to differences in the actual words instantiating the contrast (e.g. for the [ɵ] vowel, labials used [bɵs] while coronals used [dɵm] and velars used [gɵld]). They don't appear to confound any of our assessments of VOT (since with one exception, all the words had minimal pairs for voicing, and speaking rate did not interact with them), and won't be discussed further.

4.2 VOT.

Having demonstrated that our speaking rate manipulation achieved the desired results and did not differentially affect fortis or lenis stops, we next turn to our primary question: does speaking rate affect VOT for both types of stops? To assess this, we examined each type of stop separately, in a repeated measures ANOVA examining speaking rate (fast/slow) along with the secondary factors place of articulation (labial, coronal, velar) and vowel ([a], [ɔ], [ɵ]).

Our first analysis examined the *lenis* stops (Figure 2). Results were straightforward. In fast speech, lenis stops showed significantly less prevoicing (M=−78.5 ms; SD=19.5 ms) than in slow speech (M=−107.9; SD=18.0; $F(1,7)=43.1$, $p=0.0003$). This did not interact with place ($F(2,14)=1.9$, $p=0.19$) or

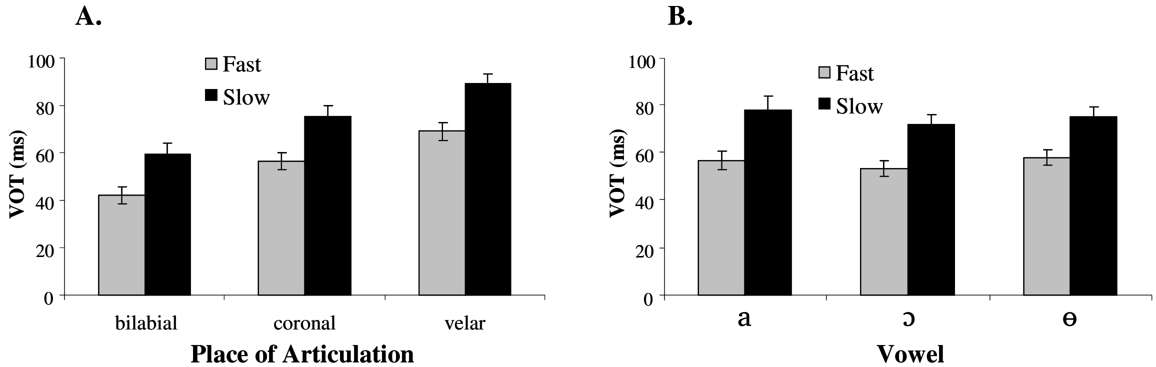

Figure 3: Effect of speaking rate on VOT for lenis stops A) broken down by place of articulation; B) broken down by vowel.

vowel (F(2,14)=1.0, p=0.37) nor did it participate in a three-way interaction (F<1). Thus speaking rate appears to have a large effect (over 30 ms) on VOT in prevoiced stops in CS Swedish.

This analysis also found a handful of predicted effects of other factors on VOT in these stops. Place of articulation (see Figure 2A) significantly affected VOT (F(2,14)=51.4, p<.0001), with bilabials having much more negative VOTs (M=−111.7, SD=15.4) than coronals (M=−86.3, SD=21.5; F(1,7)=37.9, p<.0001), which in turn had more negative VOTs than velars (M=−81.3, SD=17.8; F(1,7)=8.9, p=.02). Vowel also had a significant effect on VOT (F(2,14)=3.7, p=0.05), though it was much smaller than those of speaking rate and place (Figure 2B). This was driven by the fact that VOT in the stops preceding [ɵ] (M=−96.4 ms SD=15.6) was slightly more negative than those before both [ɔ] (M=−90.9 ms, SD=19.8, T(7)=2.4, p=.046) and [a] (M=−92.0, SD=18.5, T(7)=1.86, p=.11). Finally, the place x vowel interaction was also significant (F(4,28)=2.7, p=0.048). This was driven by the fact that coronals differed from velars in the [a] contexts ([a]: T(7)=4.3, p=.003), but not in the other two contexts ([ɔ]: T<1; [ɵ]: T<1). Labials differed in all three ([a]: T(7)=4.5, p=.002; [ɔ]: T(7)=−5.8, p=.0006; [ɵ]: T(7)=5.7, p=.0002).

Next we examined *fortis* stops (Figure 3). Here we again found a main effect of speaking rate (F(1,7)=113.6, p<.0001). VOTs in fast speech averaged 55.8 ms (SD=9.6), while those in slow speech averaged 74.5 ms (SD=12.3). As with lenis stops, this did not interact with place (F<1), vowel (F<1), or participate in a three-way interaction (F<1). Thus, there is clear evidence that both fortis and lenis stops are affected by speaking rate.

As before, the effect of place was significant (Figure 3A) with bilabials shorter than coronals (T(7)=4.8, p=.0019) and coronals shorter than velars (T(7)=19.8, p<.0001). This time, however, vowel was not significant (F(2,14)=1.7, p=0.22), although it did interact with place (F(4,28)=3.5, p=0.019). This was due to the fact that bilabials did not differ from coronals in the [a] context (T(7)=1.3, p=.24), while they did in the other contexts ([ɔ]: T(7)=5.3, p=.001; [ɵ]: T(7)=3.9, p=.006). Coronals differed from velars in all three vowel contexts ([a]: T(7)=5.6, p=.0007; [ɔ]: T(7)=5.0, p=.0015; [ɵ]: T(7)=3.8, p=.007).

Thus, these analyses of VOT support an effect of speaking rate on VOTs for both fortis and lenis stops that does not interact with any other factor. They also confirm the predicted effects of place of articulation and hint at some smaller effects of vowel on VOT.

Our final analyses were more descriptive in nature, and were intended to flesh out these effects. We first analyzed the distribution of tokens to better describe our main effect and to verify that the shifts were not due to outliers, but rather arise from a shift in the prototype values. Next, we examined the correlation between vowel length and VOT to determine if this effect was a continuous function of speaking rate.

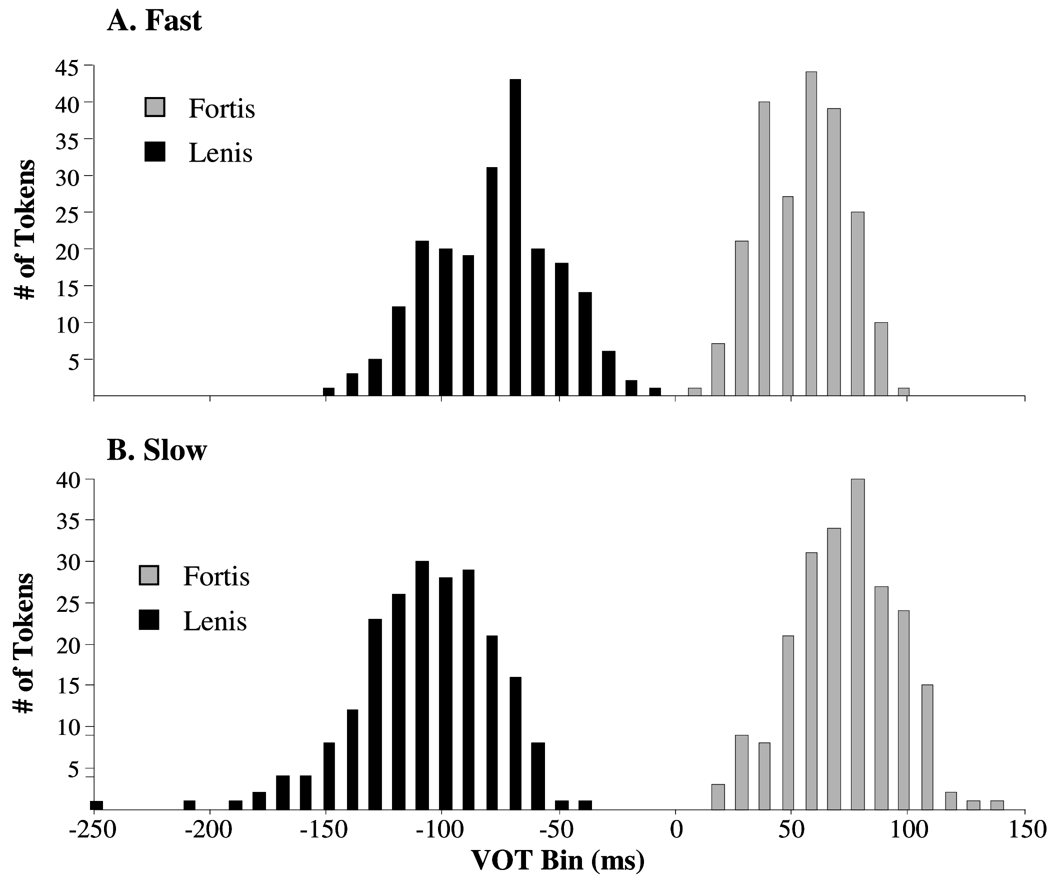

Figure 4: Distribution of VOTs in the slow (Panel A) and fast (Panel B) conditions. Bins are centered at 0, 10, 20 ms (etc).

4.3 Distribution of VOTs

To conduct our analysis of the distribution, we first binned the data into 10 ms increments of VOT centered at 0, 10, 20 etc. We then counted the number of VOTs in each of the two speaking rate conditions that were observed. The data is shown in Figure 4A for fast speech and 4B for slow. This reveals a number of important facts about this dataset. First, there was no overlap between fortis and lenis stops in either condition – that is, the fortis stop with the lowest VOT across all subjects was higher than the lenis stop with the highest. Even in fast speech, there is literally no overlap in the VOTs of fortis and lenis stops in CS Swedish. Thus, the larger difference between stop-types in slow speech does not appear to be necessary to improve the contrast – there was never any ambiguity to begin with.

Second, the entire distribution of tokens changes in response to speaking rate. The modal value of a fortis stop shifts from 60 ms in fast speech to 80 ms in slow; likewise, for lenis stops the mode shifts from −70 ms in fast speech to −110 in slow. Similarly, the central edges of the distribution also expand—for fast speech, the region at which no tokens were observed is from −10 to 10 ms, while for slow speech it extends from −30 to 20. Thus, these effects do not appear to be driven by outliers. Rather, speaking rate affects the entire distribution of tokens.

4.4 Relationship between VOT and word duration

The final analysis examined the relationship between speaking rate and VOT as a more continuous measure. Figure 5 shows the relationship between word duration (across both conditions) and VOT as a

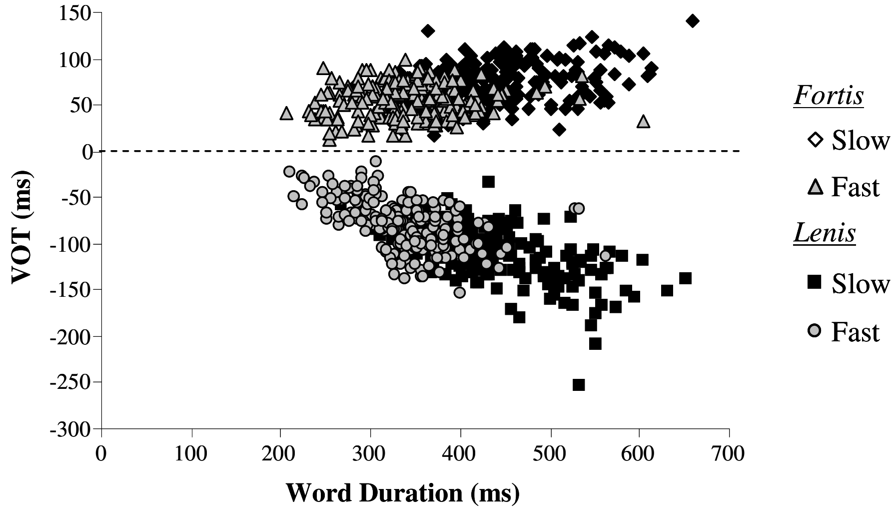

Figure 5: The relationship between word duration (speaking rate) and VOT for both fortis and lenis stops in both rate conditions.

scatter plot of the raw data. Again, this emphasizes the lack of ambiguity in CS Swedish, showing no tokens of lenis stops with VOTs greater than 0 (and vice versa for fortis). It is also clear that the relationship between VOT and speaking rate is largely due to the word duration. Within stop-voicing, the two rate conditions showed considerable overlap and a similar relationship between word duration and VOT. Finally, it appears that the relationship between speaking rate and VOT may be slightly stronger for prevoiced than voiceless sounds.

To verify this, a linear regression analysis[4] examined the effect of word-duration, rate condition, and their interaction on VOT. Crucially, if the slopes of the relationship between word duration and VOT were the same in the two conditions, we should find no interaction between rate-condition and word-duration (which would indicate that the effect of word-duration was different in the two VOTs).

The first model examined lenis stops. In the first step, seven subject codes were added to account for subject variance, and together they accounted for 27% of the variance ($F_{change}(7,424)=22.4$, $p<.001$). In the second step, word duration accounted for an additional 33.7% of the variance ($F_{change}(1,423)=187.8$, $p<.0001$). Individually, word duration showed a negative relationship with VOT – as words lengthened, VOTs shortened by about .267 ms for each ms of word duration. On the third step, Rate condition accounted for very little variance ($R^2_{change}=.007$), but was significant ($F_{change}(1,422)=7.2$, $p=.008$) with a difference of about 7 ms between the two conditions. On the final step, the interaction accounted for no new variance ($F_{change}<1$). Thus, for lenis stops, there is a linear relationship between word duration and VOT, and the effect of manipulating speaking rate appears to be largely due to its effect on word duration.

This analysis was repeated for the fortis stops. In the first step, subject accounted for 20.5% of the variance ($F_{change}(7,423)=15.5$, $p<.001$). In the second step, word duration accounted for an additional 10.7% of the variance ($F_{change}(1,422)=65.8$, $p<.0001$). This time, word duration had a positive relationship with VOT–as words lengthened, VOTs also lengthened, by about .10 ms for each ms of word duration. On the third step, rate-condition accounted for 7.3% of the variance, ($F_{change}(1,421)=49.9$, $p<.001$), with a difference of about 16 ms between the two conditions. Thus, subjects' conscious manipulation of rate may have had more of an explicit effect on positive VOTs. On the final step, the interaction was significant ($F_{change}(1,420)=6.4$, $p=.012$) but accounted for a very small amount of the variance($R^2_{change}=.009$) compared to the other two factors. Thus, for fortis stops, speakers may have been doing something slightly different in the fast condition than the slow (over and above the

---

4 *Regressions were run hierarchically using a subject code? for repeated measures. This is known to misestimate error-variance and degrees of freedom, and Cohen and Cohen (1983, p.432) recommend an adjustment to F and df to compensate. For simplicity, we report unadjusted statistics; all contrasts were significant using their adjustment as well.*

effects driven by word length).

Figure 5 and the slope estimates in the previous analyses suggest the relationship between word duration and VOT may differ across the two stop types. To examine this, a final regression asked if the relationship between the magnitude (absolute value) of VOT and word duration was different in fortis and lenis stops. As before, the analysis was conducted hierarchically, although this time the entire dataset was used and VOT was converted to its absolute value (so that the slopes for fortis and lenis stops would both be positive). As in the previous analyses, subjects accounted for a substantial amount of the variance (R^2_{change}=.152; $F_{change}(7,855)$=21.8, p<.0001). In the second step, underlying voicing accounted for an additional 20.5% of the variance ($F_{change}(1,854)$= 272.8, p<.0001): fortis stops had absolute VOTs that were about 28 ms shorter than lenis. That is, fortis stops did not have as much aspiration as lenis stops had prevoicing. In the third step, syllable duration was positively correlated with VOT (R^2_{change}=.174; $F_{change}(1,853)$=315.9, p<.0001) with an average slope of +.179 ms of VOT for each ms of word duration. Finally, the interaction of voicing and word duration was highly significant (R^2_{change}=.045; $F_{change}(1,852)$=90.4, p<.0001) in that fortis stops had shallower slopes than lenis. Thus, the relationship between word duration and VOT, while positive for both stop types (B_{fortis}=0.10; B_{lenis}=0.276), appears to differ between them. This difference is similar to findings by Magloire and Green (1999), who also report a larger slope for lenis stops. They report slopes of .27 and .23 ms VOT/ms syllable duration for fortis stops in English and bilingual Spanish/English speakers, and slopes of −.45, and −.73 for Spanish speakers and bilinguals. Across both studies, the prevoiced stops appear much more susceptible to rate changes than aspirated stops.

This suggests that these effects on VOT are linearly related to rate (word duration), and may be an automatic consequence of timing considerations during articulation (modulated, of course, by the fact that prevoiced and aspirated VOTs are articulated differently). In fact, our slopes correspond fairly closely to those reported in the literature. For fortis sounds, we report a slope of .10 ms VOT/ms word length; Pind (1995) found a slope of .15 ms VOT/ms word length for aspirated stops in Icelandic; and Magloire and Green (1999) report a slope of .27 for English. For lenis stops, we found a slope of −.27ms VOT/ms word length; Magloire and Green report a slope of −.45 for Spanish speakers[5]. The correspondence between studies suggests that this pattern of lengthening within a stop-type may be primarily articulatory and may be independent of the language (though clearly work on more languages is needed).

As a whole, then, the lengthening appears to be an automatic consequence of articulatory timing, with differences as a function of stop-type, but perhaps not as a function of language once we control for that. However, more importantly, the lack of effects in short-lag stops in other languages suggests that it is a consequence of timing in the production of a specific feature ([sg] or [voice]), not timing in general.

5.0 General Discussion

Our results demonstrate that in CS Swedish, VOTs in both fortis and lenis stops change as a function of speaking rate. This effect was a continuous function of word length and was seen both in the mean VOTs in each category, and also throughout the category. Interestingly, this expansion in slow speech was observed despite the fact that there were literally no ambiguous tokens in this data set.

This has implications for both the phonetics of speaking rate and hyperarticuation, and for the phonology of CS Swedish (and in turn, phonology in general). We will discuss each in turn.

5.1 Speaking Rate.

---

[5] *Some of these differences may be due to the fact that we used a repeated-measures regression (partialling out the effect of speaker prior to examining rate), while Magloire & Green (1999) and Pind (1995) did not appear to do this.*

Table 3: Mean VOTs of prevoiced and/or long-lag stops reported as a function of speaking rate in several studies.

Study	Language	Stop Type	Fast	Slow	Difference
Magloire & Greene (1999)	Spanish (Mono)	Prevoiced	–45.8	–69.2	23.4
	English (Mono)	Long Lag	31.5	58.4	26.9
Kessinger & Blumstein (1997)	Thai	Prevoiced	–42	–69	27
		Long Lag	52.5	80	27.5
	French	Prevoiced	–82.5	–110	27.5
	English	Long Lag	79	107.5	28.5
Allen & Miller (1999)	English	Long Lag	48.6	78.3	29.7
Present Study	CS Swedish	Prevoiced	–78.5	–107.9	29.4
		Long Lag	55.8	74.5	18.7

CS Swedish represents an interesting language with respect to speaking rate. Even when speakers are speaking fast, there is no overlap in voicing – we found no voiceless tokens with VOTs below 0 ms, and no voiced tokens with VOTS above 0 ms. If speakers then are attempting to create "just enough" contrast to be understood (à la H&H theory: Lindblom, 1990), further manipulations of VOT in the context of slow speech are not necessary. The fact that we see such effects, then, suggests that they are the consequence of more general articulatory dynamics, not audience design.

There is some evidence that speaking rate manipulations are independent of the so-called "clear speech" mode that has been typically used to study such factors (Krause & Braida, 2002). However, a wealth of studies suggest that at the level of individual segments, cues like formant frequencies and VOTs show similar modifications due to clear-speech, slow-speech and prosodic strengthening (Bradlow & Smiljanic, 2009), modifications described as hyperarticulation.

Swedish thus presents an interesting case in which to determine whether hyperarticulation is driven by the listeners' needs for discriminating contrasts, or is more a by-product of production. Given the lack of ambiguity in even fast speech, our data suggest that such modification may not be for audience design, but rather is a by-product of the production system.

Indeed, Pind (1999) shows similar effects on vowel length in Icelandic (which has long and short vowels). He shows that a secondary cue for vowel quantity (F1) is present for [ɛ] but not [a]. From an H&H perspective, then, listeners should manipulate duration more in the context of [a], where it is the only cue, than in [ɛ]. His data do not support this: speakers modify duration equally in both, suggesting that speakers' use of durational cues may be relatively independent of listener considerations.

This is analogous to the available cross-linguistic data on VOT. Table 3 shows the mean VOT values for the marked category in several languages (Pind, 1995, and Miller, Green & Reeves, 1986, do not report means). While, of course, such data must be interpreted with caution given the likely variability in how subjects interpret instructions to speak slow or fast, across studies, we see a remarkably close correspondence. On average both long-lag and prevoiced stops move between 25-30 ms of VOT away from 0 ms. The present study found movement of –29.4 ms for prevoiced stops and +18.7 for long-lag stops, numbers that are quite in line with these values. Importantly, this results in a net gain of 48.1 ms of contrast, far more than seen in any other language. Thus, as with Pind's (1999) study of vowel quality, speakers appear to be moving their VOTs more than is necessary to obtain contrast. These facts, coupled with the close fit between the amount of movement in each category across languages, are more suggestive of an articulatory account.

Such findings cannot entirely rule out the role of audience design in speech production. It's possible that temporal cues are largely based on articulatory dynamics, but speakers do modify spectral cues for the benefit of the listener. It is also possible that voicing is too easy and such considerations play a bigger role in the articulation of contrasts that are more overlapping and/or affected by context, contrasts like vowels (e.g. Hillenbrand, Getty, Clark & Wheeler, 1995) or fricatives (e.g. McMurray & Jongman, submitted).

Nonetheless, these VOT effects appear to be motivated by articulatory dynamics, not listener concerns. However, these dynamics are not divorced from the phonology of the language, nor from speakers' intentional speech modes. It is not the case that the whole articulatory system simply speeds up or slows down—otherwise we should see larger effects on short-lag stops. Rather, the effect of speaking rate is selective for particular stops. The simplest explanation for this is that it is the particular phonetic gesture specified by the language that slows down. The dependence of the particular gesture on rate is supported by the fact that, individually, both the prevoiced and aspirated stops in CS Swedish slow lengthen by roughly the same amount as those same categories in other languages. In a sense, regardless of listener needs, speakers' intentions to speak slowly or more clearly have consequences on the articulation, but the specific consequences are driven by the gestures specified by the language. This makes a strong case for privativity, the idea that languages specify dimensions like voicing by the presence or absence of a particular feature.

5.2 Phonological Specification

We now consider what our results show about phonological feature specifications in CS Swedish. There are two possibilities: features are binary, or features are privative.

5.2.1 Binary Features.

The binary feature [voice] can be used to capture the two-way contrasts of French and English (with different phonetic implementation rules, as per Keating, 1984). However, with only [±voice], we can only describe a two-way laryngeal contrast, not a three-way contrast as found in Thai. To capture a three-way contrast, we apparently need both [±voice] and [±sg].

What, then, are the features for English? Adding the binary feature of [±sg] means that there are three possible phonological representations for a two-way contrast such as in English:

(1) *Binary feature representations for English stops*

	Lenis stops	Fortis stops
a.	[−voice, −sg]	[−voice, +sg]
b.	[+voice, −sg]	[−voice, +sg]
c.	[+voice, −sg]	[−voice, −sg]

Regardless of which of these three representations is assumed, in word-initial position, the phonetic cues for the stops would be the same: the lenis stops would be short-lag VOT and the fortis stops would be long-lag VOT (aspirated). Similarly, in CS Swedish, the same three possibilities shown in (1) above for English are available to represent the two types of stops. But in CS Swedish, the phonetic implementation for these features would have to be different: the cue for the lenis stops in initial position would be prevoicing and the phonetic cue for the fortis stops would be aspiration.

This gives rise to two competing interpretations. First, the choice of representations for a two-way contrast is entirely free, but the mapping from phonological representation is entirely arbitrary (e.g., [+voice,−sg] is used for both English and CS Swedish, realized as short-lag VOT in English, but with prevoicing in CS Swedish). Alternatively, the choice of representations is rigidly constrained by the phonetics. If there is aspiration in the phonetic realization, stops must be [−voice, +sg], if there is prevoicing, stops must be [+voice, −sg], and short-lag stops are necessarily [−voice, −sg].

This latter view of specification with binary features brings us to a position which is virtually indistinguishable from a privative feature account we shall describe next. That is, aspirated stops must be [+sg] equivalently (privative) [sg] and prevoiced stops must be [+voice], equivalently (privative) [voice]. This raises the question, then, of what binary (as opposed to privative) features buy for us. More importantly, if features are binary (and both values are specified equally), why is it always and only the case that the positive values of the features are impacted by speaking rate?

5.2.2 Privative Features

The fact that the voiceless unaspirated stops in all three languages remain unchanged provides some support for the phonologists' claim that this is the unmarked category. If we are correct in suggesting that in slower speech, VOT cues associated with specified laryngeal features are lengthened, then our results for CS Swedish suggest that both [sg] and [voice] are phonological features of CS Swedish stops, but that there are no stops unspecified for a laryngeal feature (Beckman & Ringen 2004, Ringen & Helgason 2004).

If the stops in CS Swedish are (over)specified in this way, as either [voice] or [sg], then these rate effects can be accounted for quite cleanly in the way that positive feature values are mapped onto speakers' production goals.

5.2.3 Implications for Phonology

This privative account of CS Swedish has implications for phonology more broadly. (See also Beckman, Jessen & Ringen, 2009 for arguments for overspecification of fricatives in German.) Chomsky and Halle (1968) advocate for "economical representations" in phonology, the use of the minimal features necessary to capture the contrasts of a language. CS Swedish clearly violates this principle, using two features (hence four possible values, in a binary feature system) to represent a two-way contrast. However, that appears to be the most parsimonious explanation. If we assume that [voice] is the sole feature of CS Swedish, we are left to explain why CS Swedish aspirates the other (unspecified) category of stops, but French (a [voice] vs. [∅] language) does not. Similarly, if we use just [sg] for the CS Swedish contrast, then we are left to explain why CS Swedish prevoices initial non-[sg] (i.e., unspecified) stops, but English does not. Our present data only enhance this argument. Both the fortis and lenis stops appear to be acting (phonetically) in the same way as [voice] and [sg] stops in other languages. There is no phonetic reason for assuming the prevoiced category in CS Swedish is any different from French, and similarly no reason to assume that aspirated stops in CS Swedish behave any differently from those in English.

With both [voice] and [sg] in the language, however, it is true that there are more features than are needed to make a two-way contrast. Until recently, such a system has been considered unparsimonious (but see Anderson, 1985). However, at this point, it is reasonable to ask why economy should take priority over phonetic reality. Why should we embrace the position that the first principle is economy at any cost? Why does economy have priority over faithfulness to the phonetic realization? A case in point is fricatives of many [sg] languages, for which lenis fricatives are marked with laryngeal voicing, even while stops use only [sg] (e.g. Beckman, Jessen & Ringen, 2009). Here too, an over-specified representation makes sense, and suggests that such representations are not outliers, but rather may be fairly common. Why do we attach more significance to economy of description rather than phonetic accuracy? Doing so increases abstractness for no reason other than economy, and leaves unexplained why CS Swedish and English, or CS Swedish and French, behave differently, when a clear explanation is at hand.

6. Conclusions

The influence of speaking rate on VOT in Swedish mirrors the effects that have previously been found cross-linguistically: both prevoiced and long-lag (aspirated) VOTs lengthen in slow rates of speech. Interestingly, this was found in a language that does not have an intervening short-lag category, and, as a result, no phonetic ambiguity in voicing.

These results make sense if we assume that, in slower speech, the duration of the phonetic cue for the specified feature is increased. This privative feature account, in which phonemes are defined only by the cues that are present, suggests that CS Swedish has a two-way laryngeal contrast using both [sg] and [voice], with no [∅] feature. As a result both stop types change in slower speech: both the negative VOTs and the long-lag VOTs got longer. This leads to two clear conclusions. First, this only makes sense if, in CS Swedish, both [voice] *and* [sg] are specified. While this appears to violate the principle of economy, it does so by grounding the phonological description in the phonetics of the language. Second, the effect of rate on VOT is not geared to the listener. Even in fast speech, VOTs in CS Swedish are unambiguous—slowing down (and expanding the VOTs) thus offers listeners no benefit. To the extent that this rate manipulation is representative of other hyperarticulation phenomena, it suggests that apparent changes in phonetic contrastiveness may be the natural result of articulatory dynamics, but dynamics that are intimately tied to the featural specification of the language, that is, to what speakers are intending to produce.

Acknowledgements

We'd like to thank Joanne Miller and Sheila Blumstein for providing results from their studies, Vladimir Kulikov for assistance with data analysis, and Christine Ericsdotter for her help in carrying out the experiment. This research supported by NIDCD grant 008089 to BM and a Global Scholar Award from the University of Iowa and NSF grant BCS-0742338 to CR.

References

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

Anderson, J. M. & Ewan, C. J. (1987). *Principles of Dependency Phonology*. Cambridge: Cambridge University Press.

Anderson, S. R. (1985). *Phonology in the Twentieth Century: Theories of Rules and Theories of Representations.* Chicago & London: University of Chicago Press.

Beckman, J. & Ringen, C. (2004). Contrast and redundancy in OT. In V. Chand, A. Kelleher, A. Rodriguez, & B. Schmeiser (Eds.). *Proceedings of the 23rd West Coast Conference on Formal Linguistics*, Somerville, MA: Cascadilla Press, 85-98.

Beckman, J., Jessen, M. & Ringen, C (2009). German fricatives: Coda devoicing or positional faithfulness? *Phonology 26,* 231-268.

Bradlow, A., and Smiljanic, R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass, 3(1),* 236-264.

Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English.* New York: Harper & Row.

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.

Englund, K. T. (2005) Voice onset time in infant directed speech over the first six months. *First Language 25(2),* 219-234.

Frye, R. E., McGraw Fisher, J., Coty, A., Zarella, M., Liederman, J., & Halgren, E. (2007). Linear coding of voice onset time. *Journal of Cognitive Neuroscience, 19*, 1476-1487.

Harris, J. (1994). *English Sound Structure*. Oxford: Blackwell.

Heldner, M, & Strangert, E. (2001). Temporal effects of focus in Swedish. *Journal of Phonetics*, *29(3)*, 329–361.

Helgason, P. (2002). *Preaspiration in the Nordic Languages: Synchronic and Diachronic Aspects*. Ph.D. Dissertation, Stockholm University.

Helgason, P. & Ringen, C. (2008). Voicing and aspiration in Swedish stops. *Journal of Phonetics*.

Hillenbrand, J., Getty, L, Clark, M., & Wheeler, K. (1995). Acoustic Characteristics of American English vowels." *Journal of the Acoustical Society of America, 97(5),* 3099-3111.

Honeybone, P. (2005). Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications. In M. van Oostendorp, and J. van de Weijer (Eds.) *The Internal Organization of Phonological Segments* (pp. 319-354). Berlin: Mouton de Gruyter.

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

Jessen, M. & Ringen, C. (2002). Laryngeal features in German. *Phonology* 19.2, 189-218.

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

Kessinger, R. H. & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, & 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. *Journal of the Acoustical Society of America, 112*, 2165–72.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190,* 69-72

Lisker, L. & Abramson, A. (1964). Cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384-422.

Lindblom, B. (1990). Explaining phonetic variation. A sketch of the H&H theory. In W. Hardcastle and A. Marchal (Eds.) *Speech Production and Speech Modeling* (pp. 403–439). Kluwer: Dordrecht.

Magloire, J., & Green, K. P. (1999). A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica*, 56, 158-185.

Massaro, D. W, & Cohen, M. M. (1983). Categorical or continuous speech perception: A new test. *Speech Communication, 2,* 15-35.

McMurray, B., Clayards, M., Tanenhaus, M., and Aslin, R. (2008). Tracking the timecourse of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin and Review, 15(6),* 1064-1071.

McMurray, B. & Jongman, A. (submitted). Linking fricative production and perception: Assessing invariance, exemplar and parsing the informational assumptions of models of speech perception with a naturalistic corpus of fricatives.

Mester, R. A. & Itô, J. (1989). Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language 65*, 258–293.

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

Miller, J.L., O'Rourke, T. and Volaitis, L.E. (1997). Internal structure of phonetic categories: Effects of speaking rate. *Phonetica, 54,* 121-137.

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

Ohlemiller, K. K., Jones, L. B., Heidbreder, A. F., Clark, W. W., & Miller, J. D. (1999). Voicing judgments by chinchillas trained with a reward paradigm. *Behavioural Brain Research, 14(100),* 185-195.

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

Pind, J. (1999). Speech segment durations and quantity in Icelandic. *Journal of the Acoustical Society of America, 106(2),* 1045-1053.

Ringen, C. & Helgason, P. (2004). Distinctive voice does not imply regressive assimilation: Evidence from Swedish. In P. Boersma & J. A. Cutillas Espinosa (Eds.), *International Journal of English Studies: Advances in Optimality Theory 4.2*, 53-71.

Petrova, O., Plapp, R., Ringen, C. & Szentgyörgyi, S. (2006). Voice and aspiration: Evidence from Russian, Hungarian, German, Swedish, and Turkish. *The Linguistic Review 23*, 1-35.

Pisoni, D.B. (1977). Identification and discrimination of the releative onset of two component tones: Implications for the perception of voicing in stops. *Journal of the Acoustical Society of America, 61, 1352-1361*.

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America, 106,* 1078-1083.

Smiljanic, R., & Bradlow, A. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics, 36(1),* 91-113.

Soli, S. (1983). The role of spectral cues in the discrimination of voice onset time differences. *Journal of the Acoustical Society of America, 73,* 2150-2165.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*(5), 1074-1095.

Toscano, J., McMurray, B., Dennhardt, J., and Luck, S. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science, 21(10),* 1532-1540

Volaitis, L. E. & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America 92*, 723-735.