Effects of minimal pair competitors on voice onset time and pitch accent production in South Swedish

Benjamin M. Kramer
Advisor: Dr. Jason Shaw

Submitted to the faculty of the Department of Linguistics in partial fulfillment of the requirements for the degrees of Bachelor of Arts and Master of Arts

Department of Linguistics
Yale University
April 12, 2024

Abstract

Previous findings suggest that lexical competition between minimally different words results in hyperarticulation of the contrastive features that distinguish them. I investigated whether lexical competition in minimal pairs affects production differently depending on the role and structure of the differentiating contrast in a language's phonology. As case studies, I consider the effects of minimal pair competition on the production of the pitch accent contrast and the stop voicing contrast in South Swedish; while contrastive hyperarticulation for contrasts along these dimensions has been observed in other languages, these contrasts in South Swedish have a particularly low functional load and a particularly high category distance, respectively. Results from an experimental word naming task indicate that minimal pair competition does not significantly affect voice onset time in South Swedish. For the pitch accent contrast, minimal pair competition is significantly correlated with *converged* rather than diverged accent contours. These findings are consistent with activation dynamics of phonetic planning that are sensitive to language-specific characteristics of a contrast, such as category distance and functional load.

Table of contents

Acknowledgments

I am indebted to Dr. Jason Shaw for his assistance with this work. His feedback at every step of the way has been instrumental in shaping the project from inception to completion, just as his mentorship has shaped me over my time at Yale. Over hours of brainstorming in his office and problem-solving in the phonetics laboratory, he has been incredibly generous in showing me the ropes of the trade. His enthusiasm for research and teaching is a gift to all of his students, and I am lucky to have been able to work with him.

I am also indebted to Michael Stern for his indispensable support in data analysis and for his interest in the project. His willingness to answer my many questions about coding was a great help. Manasvi Chaturvedi, too, provided crucial assistance in data analysis. I am thankful, as well, to Natalie Weber for serving as a reader for the project.

In addition, I am particularly appreciative to Axel Källenius for assistance in translating and testing the experimental materials and to Sammy Andersson for contributions to stimulus development. I must thank Julian Daniel at Yale, as well as Malin Svensson Lundmark and Arthur Holmer at Lund, for their significant contributions to participant recruitment.

Finally, I am grateful to Priscilla Ehrgood and Jem Burch for their personal support during the development of this research.

1 Introduction

Lexical competition between phonological neighbors—words that are similar to each other in phonological form—has been claimed to influence speech production in a number of seemingly disparate ways. For example, it has been argued both that words with more phonological neighbors are hyperarticulated (Munson & Solomon, 2004) and that they are more reduced (Gahl et al., 2012) relative to words with fewer phonological neighbors. Recent work has sought to shed light on such discrepancies by analyzing productions of specifically those features that differentiate words from their neighbors. For several contrastive features across the world's languages, it has been observed that minimal pairs contain more extreme productions of their differentiating features than words with no minimal pair competitor. Such so-called *contrastive hyperarticulation* (Wedel et al., 2018) in minimal pairs has been reported for both segmental (e.g., Baese-Berk & Goldrick, 2009; Buz et al., 2016; Schertz, 2013; Seyfarth et al., 2016; Wedel et al., 2018) and suprasegmental (Chow, 2020) contrasts.

I aim to determine whether competition between phonological neighbors impacts contrastive dimensions of speech differently depending on language-specific factors, such as category distance in feature space and functional load. To that end, I investigate how minimal pair competition shapes lexical-phonetic variation along the contrastive dimensions of voice onset time (VOT) and fundamental frequency (f_0) contours in South Swedish. These two contrastive dimensions in South Swedish are of interest for two reasons. First, both VOT and f_0 contours have been found to be hyperarticulated in minimal pairs in other languages: English for VOT (Baese-Berk & Goldrick, 2009; Buz et al., 2016; Schertz, 2013; Wedel et al., 2018) and Cantonese for tone (Chow, 2020). Second, the VOT contrast and the tonal contrast in South Swedish are significantly different from the English VOT contrast and the Cantonese tonal contrast. Whereas English has a VOT contrast between low positive VOT and high positive VOT (Lisker & Abramson, 1964), Swedish has a contrast between high negative VOT and high positive VOT (Beckman et al., 2011). Whereas Cantonese has a robust lexical tone contrast, the Swedish pitch accent contrast—between "accent 1" and "accent 2"—is morphophonologically conditioned and has been argued to be communicatively redundant (Elert, 1972; Riad, 1998; Roll, 2022). These contrasts in South Swedish, thus, are well-suited to provide insight into how language-specific parameters may influence the role of minimal pair competition in lexical-phonetic variation.

I collect productions of monosyllabic words beginning with voiceless stops and disyllabic words with both accent 1 and accent 2 in an online word naming task that follows the experimental design of Baese-Berk and Goldrick (2009). I compare VOT and f_0 contours between words with minimal pair competitors along these dimensions and words without such competitors. Furthermore, I investigate whether the contextual salience of minimal pair competitors contributes to online lexical-phonetic variation. My acoustical analysis includes tokens produced by 23 speakers of South Swedish.

This text is organized as follows. In the remainder of §1, I review previous studies of lexical-phonetic variation associated with minimal pairs, explore possible mechanisms behind such variation, and discuss why contrasts in VOT and f_0 in South Swedish are of theoretical interest. In §2, I present the materials and methodology of a word naming task designed to assess the effects of minimal pair competition and competitor salience on lexical variation in these contrasts, and in §3, I describe the results of acoustical analysis of this data. In §4, I consider the implications of these findings for the dynamics of speech planning. In §5, I conclude.

1.1 Minimal pairs and lexical-phonetic variation: Previous findings

Minimal pairs consist of words that are phonologically different only in the value of a single segment or feature. For example, the English words *cap* (/kæp/) and *gap* (/ɡæp/) form a minimal pair because they differ only in the primary phonological cue of voice onset time—the duration of time lag between stop release and the onset of voicing—of the initial velar stop consonant. While English "voiceless" /k/ is characterized by a relatively long VOT, English "voiced" /ɡ/ has a relatively short VOT (Lisker & Abramson, 1964); otherwise, the words are phonologically identical. In this way, *cap* and *gap* are *minimal pair competitors*.

It has been reported that English VOT values tend to be more extreme—voiceless VOT even longer and voiced VOT even shorter—in words with an opposite-voicing minimal pair competitor (e.g., *cap* vs. *gap*) relative to words without such a competitor (e.g., *catch* vs. **gatch*). This *contrastive hyperarticulation* effect (Wedel et al., 2018) has been observed in VOT minimal pairs in both experimental (Baese-Berk & Goldrick, 2009; Buz et al., 2016; Schertz, 2013) and conversational (Wedel et al., 2018) speech. Similar results have been reported for VOT of Korean lenis stops undergoing sound change (Jeong & Wedel, 2023). This observation is not limited to VOT;

hyperarticulated productions in minimal pairs have also been found for durational cues to voicing (Seyfarth et al., 2016) and vowel formants (Clopper & Tamati, 2014; Wedel et al., 2018) in English, as well as f_0 contrasts between lexical tones in Cantonese (Chow, 2020). Notably, it has been demonstrated that the existence of a minimal pair differing on the particular phonetic dimension of interest is a better predictor of hyperarticulation along that dimension than a general measure of lexical competition, such as phonological neighborhood density (Wedel et al., 2018).

Crucially, it appears that lexical-phonetic variation associated with minimal pairs is dynamic, such that the magnitude of hyperarticulation may vary online across contexts. Schertz (2013) observes amplified contrastive hyperarticulation in repetitions of words misheard as minimal pair competitors. Similarly, Baese-Berk and Goldrick (2009) observe that the amount of VOT hyperarticulation observed in English words with minimal pair competitors differs based on the salience of the competitor in the communicative context. They compared VOT across three display conditions in a word naming task:

(1) NO COMPETITOR. Participants see three words on a screen and are instructed to read aloud a target word beginning with a voiceless stop. That word has no voiced minimal pair competitor in the lexicon, and the other two words on the screen are unrelated.
(2) ABSENT COMPETITOR. Participants see three words on a screen and are instructed to read aloud a target word beginning with a voiceless stop. That word has a voiced minimal pair competitor in the lexicon, but it is not presented on the screen. Instead, the other two words on the screen are unrelated.
(3) SALIENT COMPETITOR. Participants see three words on a screen and are instructed to read aloud a target word beginning with a voiceless stop. That word has a voiced competitor in the lexicon, and it is one of the other words presented on the screen. The third word on the screen is unrelated.

They find that mean VOT in the ABSENT COMPETITOR condition is approximately 5 ms longer (i.e., more extreme) than in the NO MINIMAL PAIR COMPETITOR condition, indicating contrastive hyperarticulation within minimal pairs. Furthermore, mean VOT in the SALIENT COMPETITOR condition is 5 ms longer than in the ABSENT COMPETITOR condition, indicating a role of competitor salience in modulating lexical-phonetic variation.

Chow's (2020) finding of contrastive hyperarticulation in suprasegmental minimal pairs also indicates that competitor salience is a modulating factor in shaping variation. In a word naming task similar to that of Baese-Berk and Goldrick (2009), she finds that f_0 values of two level tones in Cantonese (tone 3 and tone 6) are more different from each other in productions of words presented alongside an opposite-tone minimal pair competitor than in productions of words presented alongside a same-tone distractor. The same hyperarticulation effect is found along the dimension of f_0 range for two rising tones; the difference in f_0 range between these two tones—tone 2 has a greater rise than tone 5—is amplified when an opposite-tone minimal pair competitor is contextually salient rather than not.

Hyperarticulation of contrastive features has even been found in pseudoword minimal pairs. VOT of word-initial voiceless stops is significantly longer in English pseudowords with a real word voiced minimal pair competitor (e.g., **cag* vs. *gag*) than in words without such a competitor (e.g., **cappy* vs **gappy*), although this effect is only found when the competitor is contextually salient (Stern & Shaw, 2023a). This finding indicates that contrastive hyperarticulation in minimal pairs can emerge online in the production of a relevant contrast, even in phonological sequences that are absent from the mental lexicon.

In summary, a growing body of acoustic studies supports the notion that dynamic contrastive hyperarticulation is a ubiquitous characteristic of minimal pairs, and I am not aware of any prior work that provides evidence to the contrary. Existing studies of lexical-phonetic variation in minimal pairs have focused on few languages and few types of contrasts, however; hyperarticulation in VOT minimal pairs, for example, has been investigated only in English and Korean, to the best of my knowledge. Might different patterns of lexical-phonetic variation emerge when viewing a wider range of contrasts in a wider range of languages? In the next subsection, I consider why this may be the case given existing proposals for the mechanism(s) behind such variation.

1.2 Possible mechanisms for competition effects on phonetics

Accounts of contrastive hyperarticulation in words with minimal pair competitors typically credit at least one of three possible mechanisms: (1) long-term feedback loops between produced and perceived episodic memories; (2) attention to the communicative needs of the listener; and (3) speaker-internal interactive activation of lexical and phonological categories in speech planning

(Baese-Berk & Goldrick, 2009; Stern & Shaw, 2023a; Wedel et al., 2018). I discuss each of these in turn.

1.2.1 Production-perception feedback loops

In an exemplar model of lexical storage, words are represented as collections of prior episodic memories, cached in fine phonetic detail (Pierrehumbert, 2002). Speakers naturally produce variation from one token to the next, so exemplar clouds consist of a distribution of both relatively hyperarticulated and relatively hypoarticulated memories from production and perception (cf. Lindblom, 1990). In this model, inclusion of a perceived episodic memory in an exemplar cloud for a word requires that the token be unambiguous; if a speaker cannot determine which word a token is an exemplar of, that token will not be stored as a word memory (Pierrehumbert, 2002; Wedel, 2006). In this view, tokens of words with minimal pair competitors are likely to be ambiguous and therefore be discarded from lexical storage; those tokens that *do* qualify for inclusion are more likely to be hyperarticulated. Over time, as exemplars of words with minimal pair competitors are chosen for production by a speaker and subsequently included in the episodic memory of their interlocutors (forming a community-scale production-perception feedback loop), lexical items will tend to be represented by an abundance of hyperarticulated exemplars.

This story of the emergence of contrastive hyperarticulation hinges on a crucial assumption regarding the nature of the contrast in question: ambiguous productions along the contrastive dimension of speech must result in failure by the listener to categorize the lexical item. If a contrast were to be communicatively redundant, such that hypoarticulated tokens are still clearly identifiable, then contrastive hyperarticulation should *not* emerge over time.

Long-term production-perception feedback within communities cannot alone account for the findings of lexical-phonetic variation observed in §1.1 (e.g., Baese-Berk & Goldrick, 2009). In order to derive the dynamic effects of competitor salience and communicative intent on the magnitude of contrastive hyperarticulation demonstrated by Baese-Berk and Goldrick (2009), Chow (2020), and Schertz (2013), the mechanism behind lexical-phonetic variation in minimal pairs must be sensitive to online listener-oriented and/or speaker-internal demands.

1.2.2 Accommodation of communication partners

In light of this evidence for at least some role of online processing in shaping observed lexical-phonetic variation, it has been proposed that contrastive hyperarticulation reflects an attempt by speakers—who have an awareness of the communicative needs of their interlocutors (Bell, 1984)—to enhance perception (Lindblom, 1990). It is well-known that speakers engage in global hyperarticulation in "clear speech" registers intended for particular audiences, such as foreigner-directed speech (Kendi & Khattab, 2019) and infant-directed speech (Marklund & Gustavsson, 2020), indicating sensitivity to perceived listener needs in speech planning. It is natural to extend this theoretical motivation to account for contrastive hyperarticulation observed in repetitions of words mistaken for a minimal pair competitor (Schertz, 2013) and in subsequent productions of contrasts that have been misperceived earlier in a communicative exchange (Buz et al., 2016). These findings suggest that contrastive hyperarticulation can serve both as a repair mechanism and as a prophylactic for communicative failure. Perhaps speakers are aware that perception of words in dense phonological neighborhoods—particularly those with minimal pairs—is difficult for listeners (Luce & Pisoni, 1998), and they proactively produce hyperarticulated tokens to decrease the risk of misperception, especially if prior misperception has occurred in the exchange. Furthermore, speakers may be aware that the risk of misperception increases when a minimal pair competitor is contextually salient, and in order to offset this risk, they deploy the perceptual prophylactic of contrastive hyperarticulation to a greater extent in these cases (cf. Baese-Berk & Goldrick, 2009; Chow, 2020).

Crucially, if lexical-phonetic variation associated with minimal pairs does serve a communicative function, it is not restricted to face-to-face communication; this effect has been found in online tasks regardless of whether the communication partner is real (Chow, 2020; Stern & Shaw, 2023a) or simulated (Buz et al., 2016; Seyfarth et al., 2016). Moreover, if contrastive hyperarticulation is a preventative measure against misperception, it is possible that the effect observed in experimental clear speech is actually amplified in natural speech, where speakers may need to deploy this measure more actively to counter the effects of reduction (Wedel et al., 2018).

Like the long-term restructuring proposal (§1.2.1), the listener-oriented accommodation proposal, relies on the assumption that words are easily confused for minimal pair competitors in perception. This may not be the case for all contrasts, though. Hyperarticulation should be less helpful for a listener in disambiguating between categories that are widely separated on their contrastive dimension (e.g., /ɒ/ and /u/ for vowel height) than in disambiguating between categories that are close in value (e.g., /o/ and /u/). Similarly, if a contrast is communicatively redundant, then avoidance of misperception may not be a priority for the speaker, who must balance the cost of conveying a message to the listener with the cost of articulating clear tokens (Lindblom, 1990).

1.2.3 Interactive activation of lexical items and phonological categories in speech planning

It is also possible that lexical-phonetic variation associated with minimal pairs may be an entirely speaker-internal phenomenon, emerging in the process of cascading interactive activation of lexical and phonological information in speech planning. This hypothesis emerges from a theory of retrieval in which activation of a target lexical item activates its constituent phonological categories, which in turn trigger the activation of *non-target* lexical items in which they are also found (Dell, 1986; Goldrick, 2006). When words in minimal pairs are activated, then, their competitors are also activated as a result of their overlapping phonological material. In order to overcome these activated non-target items, the activation of the target item must be "boosted" relative to words without the same amount of competition at the lexical level, resulting in more extreme productions of the features that differentiate the target from its competitor (Baese-Berk & Goldrick, 2009). Compensatory activation of the target item must be boosted further when activation of competitors is increased, which may occur when the competitor is contextually salient (cf. Chow, 2020). While these processes may benefit listeners, they are speaker-internal, facilitating production of intended outputs in the face of lexical competition.

The activation dynamics of speech planning have been modeled in Dynamic Field Theory (DFT; Schöner et al., 2016). In DFT models of neural activation, continuous dimensions in the real world—such as spatial location or f_0—are mapped onto a *dynamic neural field*. Activation of a movement plan is modeled as stabilization of an activation peak at a location in this field, which may develop under the influence of multiple inputs. Inputs to a field, which may represent activated or partially activated categories, take the form of Gaussian distributions centered at a particular field location. Stern and Shaw (2023a) derive contrastive hyperarticulation of English voiceless VOT in this framework by modeling peak stabilization under the influence of inputs in a VOT planning field. For words without a minimal pair competitor, only one input, associated with the voiceless target VOT, enters the field, and the peak stabilizes at the center of the input distribution. For words with a voiced competitor, however, an additional inhibitory input, associated with the voiced competitor VOT, enters the field. If the inhibitory competitor input and the excitatory target input are partially overlapping, as modeled by Stern and Shaw (2023a), the overlapping component of the target input distribution will become inhibited. This results in peak stabilization at a higher VOT value than the center of the target distribution (i.e., contrastive hyperarticulation). Similar models have derived trace effects in speech errors—convergences rather than divergences in category values (Goldrick & Blumstein, 2006)—through partially overlapping *excitatory* inputs, which drive peak stabilization at an intermediate location (Stern et al., 2022). These models are further discussed in §4.

Crucially, these DFT models of lexical-phonetic variation rely on partially overlapping distributions of categories to derive peak shifting. For categories that are widely separated, peak shifting is not predicted.

1.2.4 Summary of possible mechanisms for competition effects

I have discussed three possible mechanisms—which may not be mutually exclusive—for the effect of minimal pair competitors on productions of contrastive features. First, lexical competition may gradually lead to restructuring of exemplar distributions relative to words without competitors. Second, speakers may shift their productions of words with competitors to facilitate communication with a listener. Third, movement plans in target words may be influenced by co-active minimal pair competitors in speech planning.

Each of these approaches relies on assumptions about the nature of contrast that may not hold for all contrasts. The production-perception restructuring and perceptual monitoring accounts assume that hyperarticulation of a contrast facilitates word recognition. The speaker-internal activation account assumes that categories are close in feature space. Swedish has two phonological contrasts that challenge these assumptions and that may therefore be affected by minimal pair competition differently than has been observed in other languages. I consider these contrasts in the next section.

1.3 VOT and pitch accent in South Swedish

The effects of minimal pair competitors on contrasts in VOT and fundamental frequency have been studied in English (Baese-Berk & Goldrick, 2009; Buz et al., 2016; Stern & Shaw, 2023a;

Wedel et al., 2018) and Cantonese (Chow, 2020), respectively. Swedish has phonological contrasts along both of these dimensions, as well, but in Swedish, VOT categories in stops are widely separated (Beckman et al., 2011) and the pitch accent contrast may be marginal and communicatively redundant (Elert, 1972; Riad, 1998; Roll, 2022).

I investigate these contrasts in South Swedish (or *skånska*), a dialect which is spoken in Skåne, the southernmost province of Sweden. In this dialect, the relevant dimensions of contrast between accent 1 and accent 2 appear to be localized to the first syllable of polysyllabic words (Gårding & Lindblad, 1973), allowing for analytical comparisons between minimally differing single-peak tonal trajectories within that syllable. The following subsections review the VOT contrast and the pitch accent contrast in this dialect.

1.3.1 VOT in Swedish

Swedish stops may be voiced (/b/, /d/, /g/) or voiceless (/p/, /t/, /k/). In word-initial position, the voiced stops in Swedish (Figure 1a) are prevoiced (mean VOT of -60 to -96 ms depending on the place of articulation), such that periodic vibrations of the vocal folds occur during stop closure. Swedish voiceless stops are aspirated (VOT of 49 to 78 ms depending on the place of articulation), such that there is an interval of aperiodic noise following stop release before voicing begins (Helgason & Ringen, 2008). Overall, the category distance between voiced and voiceless stops in Swedish is approximately 120 ms (Beckman et al., 2011). This category distance is much greater in Swedish than in English (Figure 1b), where voiced stops are characterized by a short interval of aspiration (mean VOT of 0 to 20 ms depending on the place of articulation) and voiceless stops are characterized by a longer period of aspiration (mean VOT of 58 to 80 ms depending on the place of articulation) (Lisker & Abramson, 1964). In English, the voicing contrast may be better understood as a *lenis-fortis* distinction rather than a true *voiced-voiceless* distinction, as is found in Swedish (Beckman et al., 2011). It should be noted that the VOT measurements for Swedish referenced here are based on recordings by speakers of the Central Swedish dialect, although VOT in this dialect is not expected to differ dramatically from VOT in South Swedish.

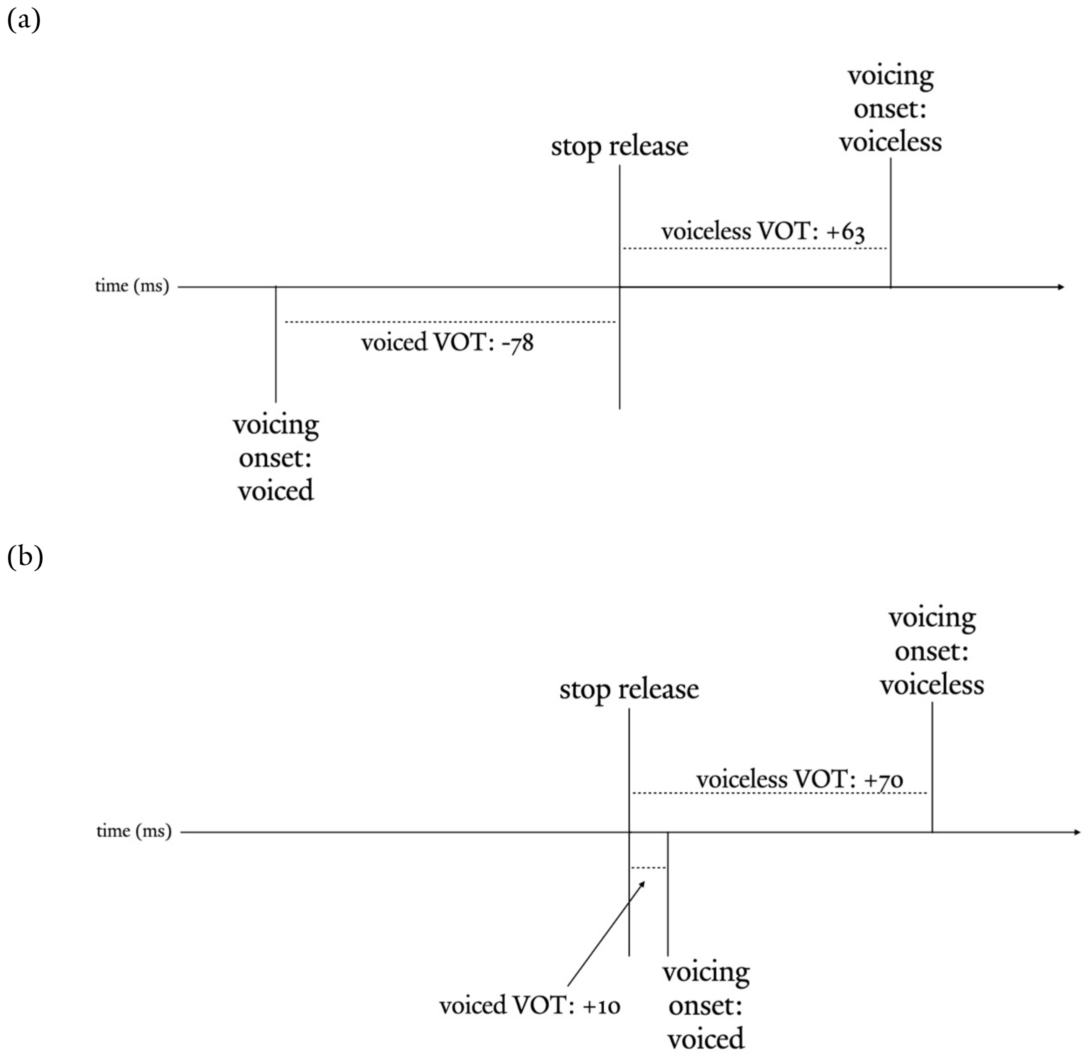


Figure 1. VOT of voiced and voiceless stops in Swedish (a) and English (b).

It is possible that this cross-linguistic difference in the distance between voiced and voiceless VOT distributions may result in different patterns of lexical-phonetic variation in VOT minimal pairs in English and Swedish. If category distance along a contrastive dimension must be small in order for activated competitor categories to interfere with activated target categories (cf. Stern & Shaw, 2023a), then it is possible that minimal pair competition will not affect VOT productions in Swedish since the category distance is so large. Similarly, assuming productions of minimal pairs are optimized so as to facilitate word recognition for the listener (cf. Lindblom, 1990), it is possible that contrastive hyperarticulation will not be observed in Swedish VOT minimal pairs, since the widely separated categories are already unlikely to be ambiguous.

1.3.2 The pitch accent contrast in South Swedish

All Swedish polysyllabic words carry one of two *pitch accent* (or *word accent*) contours: *accent 1* and *accent 2*. There is much dialectal variation in the form of accent 1 and accent 2: in all dialects, the f_0 contour of accent 1 is characterized by a single peak, whereas accent 2 has pitch peaks in two syllables in some dialects and one syllable in others (Gårding & Lindblad, 1973). South Swedish falls into the latter category. In this dialect, accent 1 is characterized by an initial high pitch that falls over the course of the first syllable (Figure 2a), and accent 2 is characterized by a pitch peak close to the ending boundary of the first syllable (Figure 2b), with a fall in the second syllable (Gårding, 1977; Gårding & Lindblad, 1973).

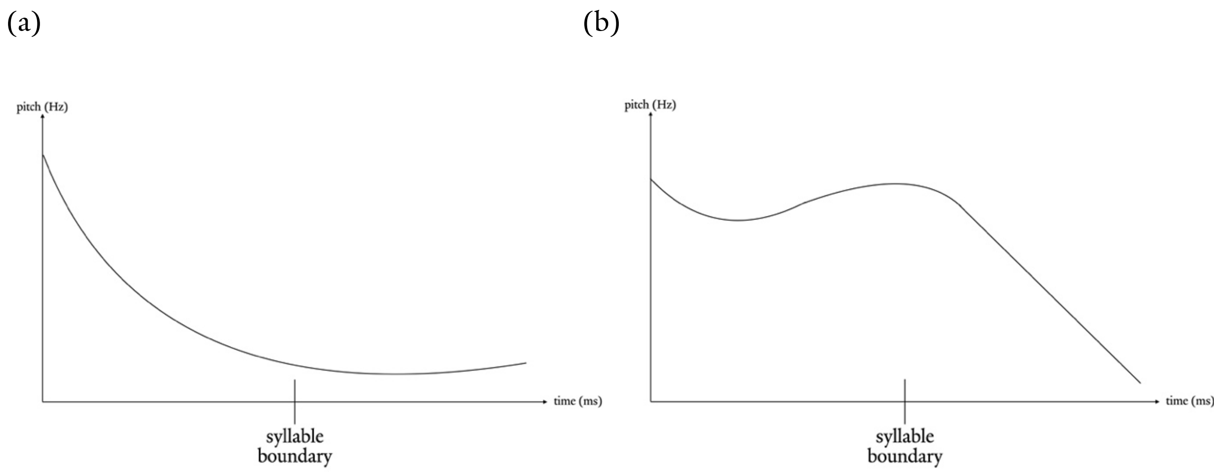


Figure 2. Schematic diagrams of accent 1 (a) versus accent 2 (b) in South Swedish.

The phonological form of accents 1 and 2 in South Swedish has been much debated. Equipollent proposals for accent specification in Swedish (Bruce, 1977; Gussenhoven, 2004; Gussenhoven & Bruce, 1999) argue that accents 1 and 2 consist of the same lexical pitch sequence, with the difference between them arising from different alignments to segmental material; in South Swedish, this pitch sequence (presumably H*L) is aligned to the beginning of the syllable in words with accent 1 and at the end of the syllable in accent 2. Haugen (1967), on the other hand, proposes that while both accent 1 and accent 2 have components consisting of intonational tones—conferred by the feature [+STRESS]—only accent 2 is also lexically marked with a privative [TONE] feature; accent 1, then, is an entirely intonational contour. Adhering to this view, Riad (2006) argues that both accent

1 and accent 2 in South Swedish consist of an intonational prominence tone (H*) followed by a L% boundary tone; in accent 2, however, a lexical L tone precedes the prominence tone, resulting in the later pitch peak observed. Conversely, it has also been argued that accent 2 consists entirely of intonational tones, whereas accent 1 contains an additional lexical tone (Lahiri et al., 2005; Wetterlin, 2007, 2010; Wetterlin et al., 2007).

Notably, the distributions of accent 1 and accent 2 are somewhat predictable. Accent 2 may only appear in polysyllabic words with a post-stress syllable; all monosyllabic words and all words with final stress carry accent 1 (Riad, 2013). Additionally, the distributions of the accents are morphologically conditioned (example (1)).

(1) Morphological conditionings of accent 1 and accent 2 in South Swedish

a. Old Norse feminine forms ending in *-a* in the singular typically bear accent 2 in both the singular and plural forms.
b. Disyllabic nouns ending in *-e* typically bear accent 2 in both the singular and plural forms.
c. Plurals of monosyllabic nouns typically bear accent 2.
d. Monomorphemic verb infinitives typically bear accent 2.
e. Monomorphemic verbs with stems ending in *-a* typically bear accent 2 in the present tense and past tense forms.
f. Monomorphemic verbs with stems ending in a consonant typically bear accent 1 in the present tense form and accent 2 in the past tense form.
g. Verb-derived agentive nouns typically bear accent 2.
h. Verbs with unstressed prefixes bear accent 1.

As a result of these prosodic and morphological conditionings, there are few pairs of words differing only in accent in Swedish. Elert (1972) lists 357 accent minimal pairs, but a good number of these include archaic second-person plural imperative forms that have entirely fallen out of use. When non-archaic minimal pairs do occur, they are often differentiable at the morphological level even if they are segmentally identical. Examples of such minimal pairs are presented in (2).

(2) Segmentally identical but morphologically distinct minimal pairs

(a)	^{1}*and-en*	^{2}*ande-n*	monosyllabic root (accent 1) vs disyllabic root (accent 2)
	duck-DEF	spirit-DEF	
	'the duck'	'the spirit'	
(b)	^{1}*knut-en*	^{2}*knut-en*	monosyllabic root (accent 1) vs past participle (accent 2)
	knot-DEF	tie-PART	
	'the knot'	'tied'	
(c)	^{1}*led-er*	^{2}*led-er*	present tense verb (accent 1) vs plural (accent 2)
	lead-PRES	joint-PL	
	'leads'	'joints'	
(d)	^{1}*panter*	^{2}*pant-er*	monomorphemic noun (accent 1) vs plural (accent 2)
	panther	deposit-PL	
	'panther'	'bottle deposits'	

Since these minimal pairs consist of words belonging to different syntactic categories – or at least different morphosyntactic categories – accent minimal pair competitors in Swedish occur in entirely non-overlapping sentential contexts, leading some to argue that the accent contrast is an "apparently useless distinction" (Roll, 2022: 2). In communication, the accent contrast appears to be redundant, as there is always a differentiating morphological or syntactic cue that can disambiguate between possible competitors – with the possible exception of ^{1}*regel* 'rule' and ^{2}*regel* 'latch'. Simply put, the contrast may serve no function to listeners. This differs from a similar word accent contrast in Norwegian, which differentiates between thousands of minimal pairs (Leira, 1998).

It is widely recognized that the traditional categories of contrast and allophony are incapable of capturing the full range of phonological relationships between sounds in human languages. The Swedish accent contrast, which is largely predictable and possibly functionless, may constitute one such "intermediate phonological relationship" (Hall, 2013). This may have interesting consequences for lexical-phonetic variation. If the accent contrast is only marginally useful to listeners in word recognition, then perhaps it will not be subject to lexical-phonetic variation associated with minimal pair competition if such variation is motivated by ambiguity. By extension, if speakers rarely must select between minimal pair competitors for inclusion in the same sentential position (as a result of differing morphosyntactic distributions), it is possible that speakers do not have practice inhibiting tonal competitors to facilitate production. This may result in different patterns of lexical-phonetic variation than have been observed for other "robust" (Hall, 2012) contrasts.

1.4 The present study

I aim to answer two crucial questions. First, what are the effects of minimal pair competitors on VOT and pitch accent productions in South Swedish? To that end, I conduct an online word naming task (cf. Buz et al., 2016) that largely replicates the methodological design of Baese-Berk and Goldrick (2009). I collect productions of word-initial voiceless stops and pitch accent contours from South Swedish speakers and compare relevant dimensions of the acoustic signal between words with minimal pair competitors along those dimensions and words without such competitors. Three outcomes are possible for each contrast. First, it may be that productions of the two contrastive categories are *more different* from each other in words with competitors than in words without competitors (contrastive hyperarticulation). Second, it may be that there is *no difference* in productions between words with competitors and words without competitors. Third, productions of the two contrastive categories may be *more similar* to each other in words with competitors than in words without competitors.

Second, I ask whether these effects are modulated by competitor salience. I compare productions of words with minimal pair competitors between conditions where the competitor is contextually salient and conditions where the competitor is not salient. There are again three possible outcomes for each contrast. First, it may be that productions of words with salient competitors are no different than productions of words with non-salient competitors. Second, it may be that productions of words with salient competitors are further shifted in feature space than productions of words with non-salient competitors. Third, it may be that productions of words with salient competitors are less shifted in feature space than productions of words with non-salient competitors.

The methodology and results of this experiment are presented in §2 and §3, respectively.

2 Methods

2.1 Participants

A call was posted in social media groups and on community forum pages relevant to Skåne for native speakers of the South Swedish '*skånska*' dialect. Speakers of this dialect are anecdotally good at self-identifying as such due to conspicuous features, such as diphthongization of long vowels and the uvular /ʁ/ or /ʀ/ (Bruce, 1970; Bruce & Engstrand, 2006). 28 participants responded to the call and completed the experiment online on the Gorilla platform (Anwyl-Irvine et al., 2019).

The data of four participants were excluded entirely for the following reasons: not completing the experimental task as instructed (n = 2) and excessive background noise (n = 2). In order to confirm that the remaining participants were speakers of the South Swedish dialect, I manually reviewed audio recordings from each participant and impressionistically assessed whether the speakers had accent 2 contours typical of the dialect. Any speaker who produced accent 2 without a pitch peak in the second syllable was considered acceptable. This manual screening process led to the exclusion of one participant, who was clearly an L2 speaker of Swedish. Additionally, participants completed a brief demographic survey at the beginning of the experiment, and only participants who indicated a place of birth and current place of residence inside the province of Skåne were included for further study. No further participants were excluded through this demographic screening process. Data from a total of 23 remaining subjects (ages 18-48, M = 28.3; 19 male, 4 female) were included in further analysis.

Participants who completed the experimental task were compensated with an electronic gift card.

2.2 Materials

Each speaker produced 48 disyllabic words, for which the f_0 contour was analyzed, and 24 monosyllabic words beginning with a voiceless stop, for which VOT was analyzed. I describe each of these sets of materials in turn.

2.2.1 Accent stimuli

Out of the 48 disyllabic words produced by each speaker, 32 words (16 with accent 1 and 16 with accent 2) had a minimal pair competitor differing only in the accent (e.g., ^{1}*anden* 'the duck'; cf. ^{2}*anden* 'the spirit'), whereas 16 (eight with accent 1 and eight with accent 2) had no such tonal competitor (e.g., ^{1}*huset* 'the house'; cf *^{2}*huset*). The target words were produced across three display conditions, in each of which the target item was presented onscreen alongside two "distractor" words:

(1) No Competitor: The target word has no accent competitor in the lexicon, and the two distractors are unrelated disyllabic words with the same accent as the target.
(2) Absent Competitor: The target word has an accent competitor in the lexicon, but it is not one of the distractors. The two distractors are unrelated disyllabic words with the same accent as the target.
(3) Salient Competitor: The target word has an accent competitor in the lexicon, and it is one of the distractors. The remaining distractor is an unrelated disyllabic word with the same accent as the target.

The 32 items with a tonal competitor were divided into two groups of 16 items each (8 with accent 1 and 8 with accent 2), and participants were randomly assigned to produce one group of items in the Absent Competitor condition and the other group in the Salient Competitor condition. Naturally, each participant produced all of the target items without a tonal competitor in the No Competitor condition.

For the sake of consistency across trials, all target items and distractors were limited to a small number of morphophonological categories for which tonal minimal pairs are found, listed in (3).

(3) Morphophonological categories of target words and distractors
- (a) Definite nouns ending in *-et* or *-en* (e.g., ^{1}*hållet* 'the direction'; ^{2}*anden* 'the spirit')
- (b) Past participles or adjectives ending in *-et* or *-en* (e.g., ^{2}*slutet* 'closed'; ^{2}*rutten* 'rotten')
- (c) Nouns ending in *-el*, *-er*, or *-ar* (e.g., ^{1}*regel* 'rule'; ^{1}*akter* 'stern (of a boat)'; ^{1}*radar* 'radar')
- (d) Plurals ending in *-er* or *-ar* (e.g., ^{2}*filmer* 'movies'; ^{2}*hundar* 'dogs')

(e) Definite plural nouns ending in *-na* (e.g., *¹bona* ‘the nests’)

(f) Verb infinitives ending in *-na* (e.g., *²bona* ‘to polish’)

(g) Present tense verb forms ending in *-er* or *-ar* (e.g., *¹leder* ‘lead(s)’; *²hoppar* ‘jump(s)’)

Because the majority of tonal minimal pairs consist of homographs (i.e., *¹regel* ‘rule’ and *²regel* ‘latch’ are written identically as <regel>), a short explanation in a smaller font size was included below each target word and distractor to disambiguate between multiple possible meanings with different pronunciations. These explanations were included for all target words and distractors, regardless of whether they had a homograph. Table 1 presents examples of target words, distractors, and explanations from each condition, alongside English translations. A full list of the accent stimuli (and when present, tonal competitors) is presented in List A (Appendix).

		Target Word	**Distractor 1**	**Distractor 2**
No Competitor	Accent 1	¹axel (kroppsdel) *shoulder* *(body part)*	¹fåren (ett får) *the sheep* *(a sheep)*	¹skriker (att skrika) *screams* *(to scream)*
	Accent 2	²ängel (väsen med vingar) *angel* *(being with wings)*	²svulten (hungrig) *starving* *(hungry)*	²vantar (en vante) *mittens* *(a mitten)*
Absent Competitor	Accent 1	¹anden (en and) *the duck* *(a duck)*	¹meter (längdmått) *meter* *(measure of length)*	¹torget (ett torg) *the city square* *(a city square)*
	Accent 2	²buren (att bära) *carried* *(to carry)*	²stolar (en stol) *chairs* *(a chair)*	²scener (en scen) *scenes* *(a scene)*
Salient Competitor	Accent 1	¹spana (ett spa) *spas* *(a spa)*	²spana (att speja) *observe* *(to spy)*	¹hajen (en haj) *the shark* *(a shark)*
	Accent 2	²tanken (en tanke) *the thought* *(a thought)*	¹tanken (en tank) *the tank* *(a tank)*	²katter (en katt) *cats* *(a cat)*

Table 1. Examples of accent stimuli. Accent subscripts are presented for reference here, but they are absent in the experimental display. Disambiguating explanations are included in parentheses.

2.2.2 VOT stimuli

In addition to the disyllabic accent stimuli, each speaker produced 24 monosyllabic CV(C)(C) words beginning with a voiceless stop, for which VOT was analyzed. Out of these 24 monosyllabic words, eight started with /p/, eight started with /t/, and eight started with /k/. 16 of these words had a minimal pair competitor differing only in the voicing of the word-initial consonant (e.g., *tolk* 'interpreter'; cf. *dolk* 'dagger'), and eight of these words did not have such a voicing competitor (e.g., *päls* 'fur'; cf. **bäls*). The target words were produced across the same three display conditions as the accent stimuli:

(1) No Competitor: The target word has no voicing competitor in the lexicon, and the two distractors are unrelated monosyllabic words beginning with /f/ or /s/.
(2) Absent Competitor: The target word has a voicing competitor in the lexicon, but it is not one of the distractors. The two distractors are unrelated monosyllabic words beginning with /f/ or /s/.
(3) Salient Competitor: The target word has a voicing competitor in the lexicon, and it is one of the distractors. The remaining distractor is an unrelated monosyllabic word beginning with /f/ or /s/.

The 16 items with a voicing competitor were divided into two groups of eight each, and participants were randomly assigned to produce one group of items in the Absent Competitor condition and the other group in the Salient Competitor condition. All of the target items without a voicing competitor were, of course, produced in the No Competitor condition by each participant.

In order to maintain consistency across all experimental trials, trials containing VOT stimuli also included an explanation in a smaller font size below the target words and distractors. Table 2 presents examples of target words, distractors, and explanations from these trials, alongside English translations. A full list of VOT target words (and when present, voicing competitors) is presented in List B (Appendix).

	Target Word	Distractor 1	Distractor 2
NO COMPETITOR	pass (resehandling) *passport* *(travel document)*	sill (fisk) *herring* *(fish)*	far (pappa) *father* *(dad)*
ABSENT COMPETITOR	tolk (översättare) *interpreter* *(translator)*	snack (prat) *chat* *(talk)*	fel (misstag) *wrong* *(mistake)*
SALIENT COMPETITOR	kolv (maskindel) *piston* *(machine part)*	golv (som man går på) *floor* *(that one walks on)*	flod (vattendrag) *river* *(body of running water)*

Table 2. Examples of VOT stimuli. Disambiguating explanations are included in parentheses.

2.2.3 Filler items

In addition to these target items, participants recorded eight filler trials in which the target word was a disyllabic word with no tonal minimal pair competitor and the two distractors constituted a tonal minimal pair (e.g., target word: ^{1}*huset*, distractors: ^{1}*modet*, ^{2}*modet*). Recordings of filler words were not analyzed.

2.3 Procedure

The word naming task was implemented in the Gorilla Experiment Builder tool (Anwyl-Irvine et al., 2019). Upon entering the online experiment in the Internet browser of their choice, participants filled out a brief demographic survey (see §2.1) and completed a microphone test to ensure that the built-in computer microphone was functional.

This experiment employed the "simulated partner paradigm" devised for a similar word naming task by Buz et al. (2016) and Seyfarth et al. (2016). Participants were told that they would be playing a communicative game with a real-life partner over the Internet, with whom they were "matched" after a 25-second waiting period. Participants were told that they had been assigned the role of a "speaker," and their partner had been assigned the role of a "listener." Both partners would see the same three words appear on their screen, but only the "speaker" (i.e., the participant) would see a box appear around one of the words. The participant's goal was to read the boxed word aloud and make their partner correctly identify which of the three words they heard. In reality, however, there was no real-life partner, and audio recordings were uploaded directly to Gorilla's data storage space. This minor deception encouraged communicative intent on the part of the speaker, which has been proposed to be essential to lexical-phonetic competition effects (Wedel et al., 2018).

On each trial, the target word and distractors appeared on the screen in a pseudorandomized order. After a delay of 5 seconds, a black box appeared around the target word. At this point, participants were instructed to read the boxed word, and the built-in computer microphone began recording. Figure 3 presents an example of the experimental display at this stage of the trial. After 5 seconds, the microphone switched off, and there was a 3-second waiting interval while the simulated partner ostensibly listened to the participant's recording and selected their answer. After this interval, a screen appeared indicating that the partner had selected an answer. The next trial began after a break of 2 seconds. A longer pause of 30 seconds was inserted after the 20th, 40th, and 60th trials, splitting the experiment into 4 blocks.

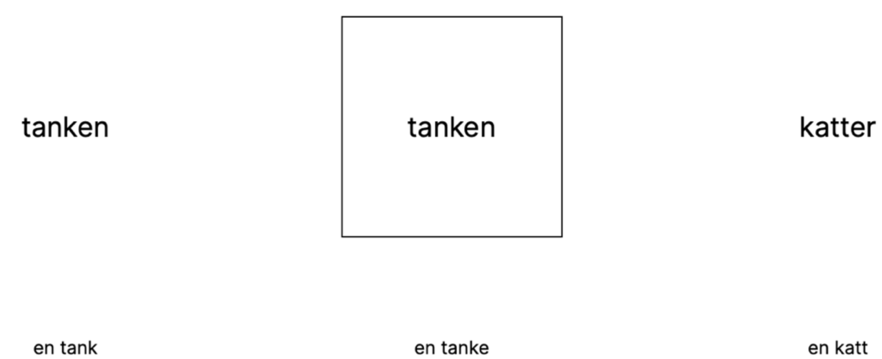


Figure 3. Experimental display.

After 5 unrelated practice trials (which were not analyzed), the accent trials and VOT trials were presented in a randomized order across all four blocks. Each participant completed 80 experimental trials (48 accent trials, 24 VOT trials, 8 filler trials). Participants produced each target word only once, and distractors did not recur across trials. The experiment lasted approximately 30 minutes.

Audio recordings were captured from participants' computer microphones and uploaded in .webm format; sampling rates varied depending on participants' computer settings.

2.4 Data processing

All non-filler audio recordings—1656 in total—were converted from .webm format to .wav format in version 3.4.2 of Audacity® (Audacity Team, 2023). I then manually inspected each audio file to determine whether the participant had produced the correct word, and if so, whether the file was of sufficiently high quality to be analyzed. A total of 23 tokens (no more than three for any participant) were excluded because I judged the participant to have mistakenly produced the word with accent 1 instead of accent 2, or vice versa. Additionally, 11 tokens were excluded because the participant mistakenly produced the word with accent 1 and then correctly repeated the word with accent 2, or vice versa; these tokens were excluded as their communicative intent (i.e., self-correction) differed from that of other tokens in the experiment. A total of 44 additional tokens were excluded for the following reasons: background noise or insufficiently high audio quality (n = 22), data storage issues (n = 11), disfluency (n = 6), empty recordings (n = 3), and yawning (n = 2). For the remaining files, I demarcated several intervals of interest in Praat (Boersma & Weenink, 2024) by hand.

For the monosyllabic items beginning with a word-initial voiceless stop, I marked voice onset time (VOT) as the interval beginning with the initial stop release burst and ending with the first periodic oscillation in the waveform. For the disyllabic items, I marked (a) the duration of the first vowel (henceforth, the *V interval*), measured as the time during which vocalic formants were visible, and (b) the duration of the pitch-supporting interval within the first syllable, excluding onsets (henceforth, the *V(C) interval*). This interval was defined as follows. For syllables without a coda consonant or with a non-sonorant coda consonant (i.e., for which the duration of the vowel is the only interval in the first syllable that supports pitch), the V(C) interval was equivalent to the V interval. For syllables with a sonorant (i.e., pitch-supporting) coda consonant, however, the V(C) interval included this consonant. If the sonorant coda was a geminate consonant, I demarcated the ending boundary of the V(C) interval halfway through this consonant in order to reflect its ambisyllabicity (Bannert, 1998). For any given word, then, the V(C) interval represents the portion of the rhyme of the first syllable that supports a pitch contour. I extracted the following measurements of interest (Figure 4) from these intervals using a custom Praat script:[1]

[1] This script was adapted from existing code in collaboration with Jason Shaw.

1. The time of the maximum f_0 in the V(C) interval as a percentage of vowel duration (*peak time*)
2. The value of the maximum f_0 in the V(C) interval (*maximum pitch*)
3. The time of the minimum f_0 in the V(C) interval as a percentage of vowel duration (*valley time*)
4. The value of the maximum f_0 in the V(C) interval (*minimum pitch*)
5. The f_0 value at the starting point of the vowel (*vowel starting pitch*)
6. The f_0 value at the ending point of the V(C) interval (*syllable ending pitch*)
7. The difference between the syllable ending pitch and the vowel starting pitch (*syllable boundary pitch difference*)
8. The duration of the V interval (*vowel duration*)

54 disyllabic tokens were excluded because f_0 in the first syllable could not be automatically detected in Praat. Out of these, 23 were excluded because f_0 was obscured by creaky voice.

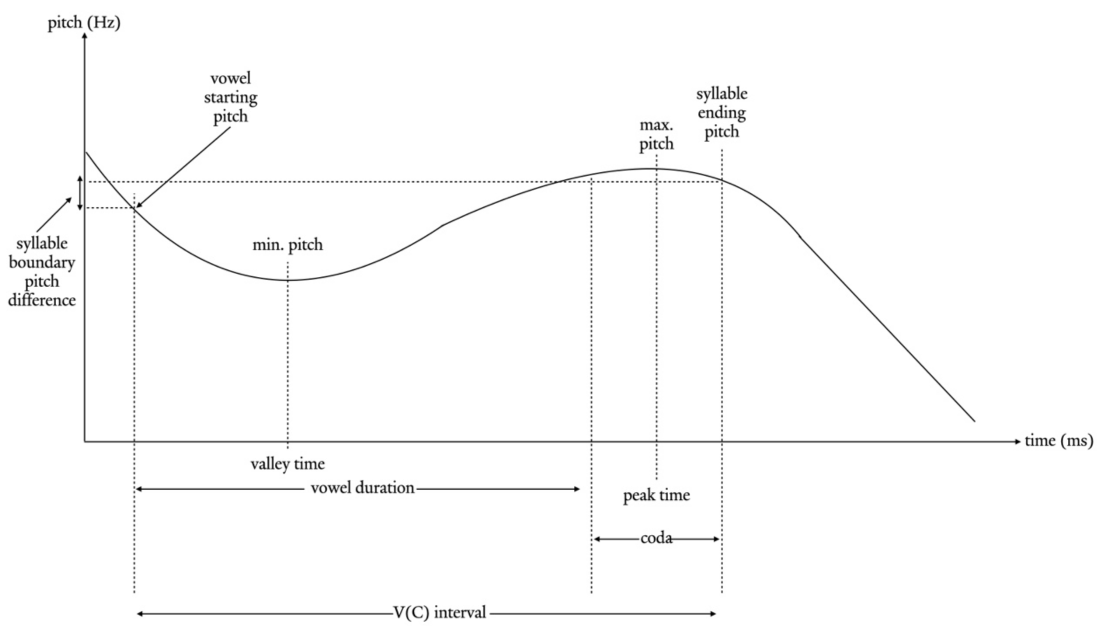


Figure 4. Schematic diagram of the measurements of interest, demarcated on a typical accent 2 contour.

3 Results

After manual exclusions (see §2.4), 1018 word-initial voiceless tokens (92.21% of total) and 532 disyllabic tokens (96.38% of total) entered into analysis. I present the results of analysis of the VOT stimuli and the accent stimuli in turn. All linear mixed effects analyses were conducted using the *lme4* package (Bates et al., 2015) in R (R Core Team, 2024).

3.1 VOT analysis

Mean VOT is 73.34 ms in the NO COMPETITOR condition, 75.12 ms in the ABSENT COMPETITOR condition, and 74.28 ms in the SALIENT COMPETITOR condition (Figure 5). It should be noted that these mean VOT values are somewhat higher than those noted for Central Swedish voiceless stops in word-initial position by Helgason and Ringen (2008). This is discussed further in §4.

A linear mixed effects analysis was employed to assess the effect of experimental condition on VOT. The model included a fixed effect of condition and a random intercept for subject. Condition was treatment coded with the NO COMPETITOR condition as the reference level. Using likelihood ratio tests, I compared this model against a baseline model with no fixed effect of condition. Table 3 exhibits the results of this model comparison. Experimental condition does not significantly improve model fit.

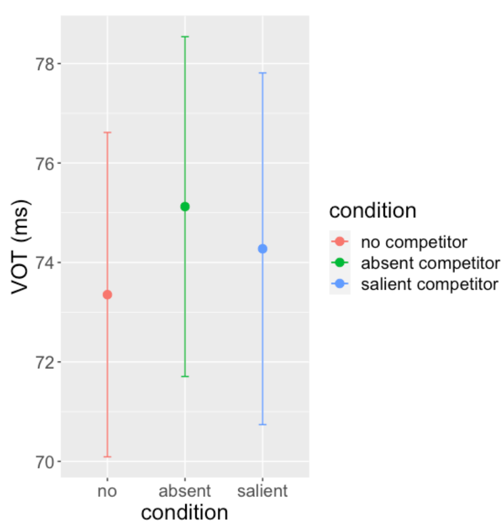


Figure 5. Mean VOT (ms) of word-initial voiceless stops with 95% confidence intervals by condition.

Model	No. of Parameters	AIC	BIC	logLik	Deviance	χ^2	df	p
baseline	3	4706.3	4719.1	-2350.2	4700.3			
+ condition	5	4809.4	4730.8	-2349.7	4699.4	0.9314	2	0.63

Table 3. Model comparison: VOT. Baseline model (with only a random intercept of subject) vs. model with a fixed effect of condition.

3.2 Accent analysis

Table 4 and Figure 6 present means for the eight measures of interest (§2.4) in words with accent 1 and words with accent 2. All absolute pitch measurements were *z*-scored by subject.

Measure	Mean Accent 1	Mean Accent 2
peak time (% of vowel duration)	18.36	67.44
maximum pitch (Hz, *z*-scored by subject)	0.06	-0.06
valley time (% of vowel duration)	83.02	46.03
minimum pitch (Hz, *z*-scored by subject)	-0.35	0.36
syllable boundary pitch difference (Hz)	-30.38	-2.60
vowel starting pitch (Hz, *z*-scored by subject)	0.30	-0.30
syllable ending pitch (Hz, *z*-scored by subject)	-0.42	0.36
vowel duration (ms)	123.00	147.78

Table 4. Mean values for the accent measures of interest, separated by accent.

I conducted a linear mixed effects analysis investigating the effect of accent on each of these measures. For each measure, I compared a model with a fixed effect of accent and a random intercept of subject against a baseline model with no fixed effect of accent. Accent was treatment coded with accent 1 as the reference level. Raw (as opposed to *z*-scored) absolute pitch measurements were employed. Table 5 presents model estimates and the result of the model comparison for each measure. Accent 1 and accent 2 are significantly different along all eight measures. Compared to accent 1, accent 2 has significantly later peak time (estimate = 42.54% of vowel duration), lower maximum pitch (estimate = -2.22 Hz), earlier valley time (estimate = -36.98% of vowel duration), higher minimum pitch (estimate = 9.58 Hz), less negative syllable boundary pitch difference (estimate = 26.88 Hz), lower vowel starting pitch (estimate = -10.65 Hz), higher syllable ending pitch (estimate = 14.42 Hz), and longer vowel duration (estimate = 24.50 ms).

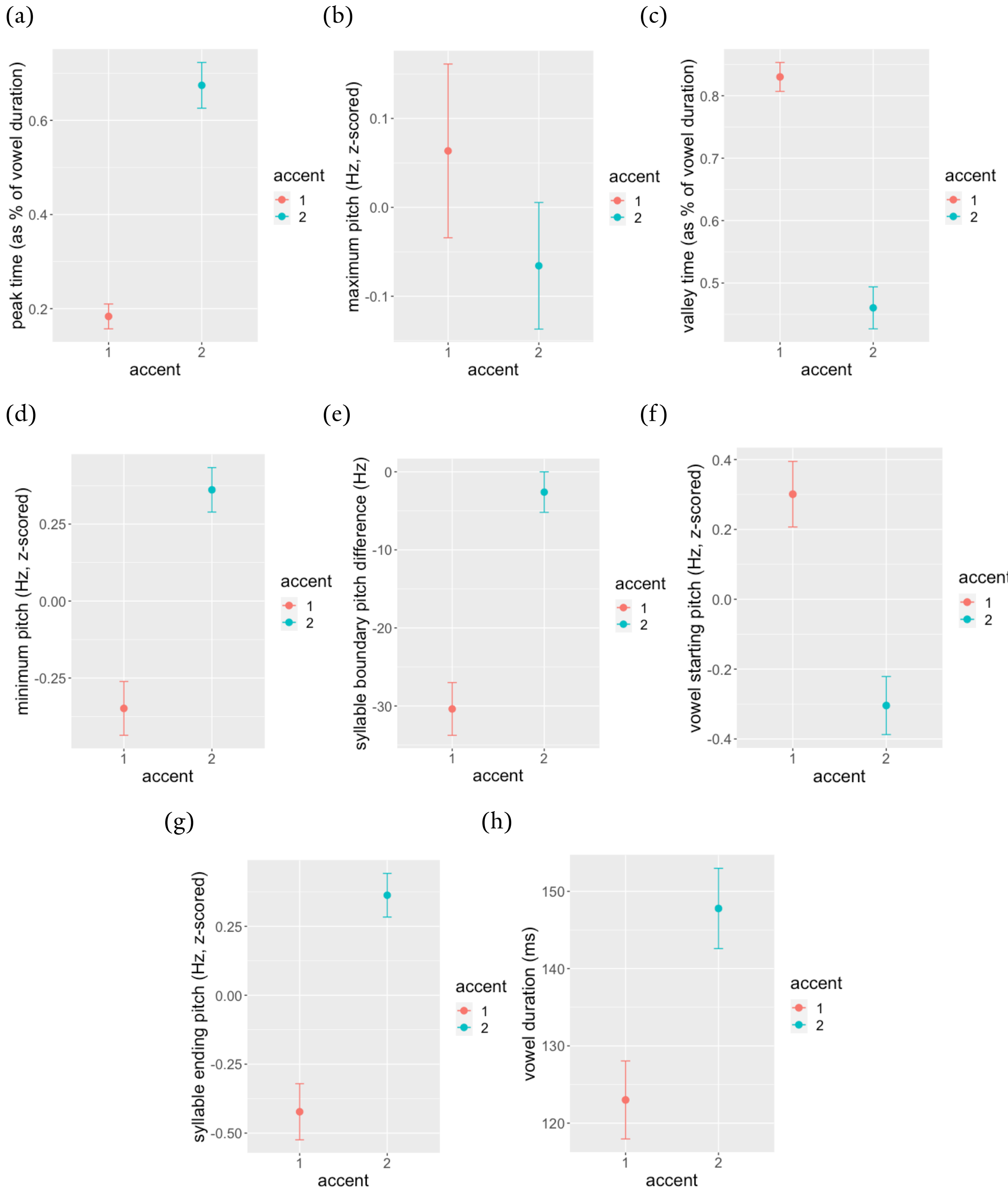


Figure 6. Means with 95% confidence intervals by accent for the following measures: (a) peak time, (b) maximum pitch, (c) valley time, (d) minimum pitch, (e) syllable boundary pitch difference, (f) vowel starting pitch, (g) syllable ending pitch, (h) vowel duration.

Measure	Model Estimate (Accent 2)	χ^2	df	p
peak time (% of vowel duration)	42.54	296.64	1	<0.001 ***
maximum pitch (Hz)	-2.22	4.23	1	0.039 *
valley time (% of vowel duration)	-36.98	295.59	1	<0.001 ***
minimum pitch (Hz)	9.58	93.97	1	<0.001 ***
syllable boundary pitch difference (Hz)	26.88	152.89	1	<0.001 ***
vowel starting pitch (Hz)	-10.65	84.21	1	<0.001 ***
syllable ending pitch (Hz)	14.42	90.46	1	<0.001 ***
vowel duration (ms)	24.50	48.533	1	<0.001 ***

Table 5. For each accent measure, an experimental model a fixed effect of accent and a random intercept of subject was compared against a baseline model with no fixed effect. Experimental model estimates and results of experimental-baseline model comparisons for each measure are presented.

Next, I explored variation in each of these measures across the three experimental conditions: No Competitor, Absent Competitor, and Salient Competitor. Table 5 and Figure 7 present means for the eight measures by condition for both accents. Since it was unknown whether the two accents should be affected similarly by condition, words with accent 1 and accent 2 were analyzed separately.

For each measure, I compared a linear mixed effects model with a fixed effect of condition and a random intercept for subject against a reduced model with no fixed effect. Analyses were conducted separately for accent 1 and accent 2, with raw (not z-scored) measurements. The No Competitor condition was treatment coded as the reference level. Table 6 presents model estimates and the results of the model comparison for each measure. I briefly summarize the results for each measure in turn.

Measure	Accent	No Competitor	Absent Competitor	Salient Competitor
peak time	1	14.50	15.62	24.65
(% of vowel duration)	2	65.44	66.79	70.44
maximum pitch	1	0.28	-0.06	-0.06
(Hz, z-scored)	2	-0.00	-0.06	-0.14
valley time	1	88.26	85.69	75.35
(% of vowel duration)	2	41.24	51.18	45.97
minimum pitch	1	-0.34	-0.33	-0.37
(Hz, z-scored)	2	0.48	0.27	0.32
syllable boundary	1	-37.71	-28.67	-25.07
pitch difference (Hz)	2	-0.81	-4.17	-2.83
vowel starting pitch	1	0.67	0.21	0.02
(Hz, z-scored)	2	-0.29	-0.27	-0.36
syllable ending pitch	1	-0.47	-0.45	-0.35
(Hz, z-scored)	2	0.38	0.31	0.41
vowel duration (ms)	1	102.75	134.07	134.57
	2	153.76	147.08	141.69

Table 5. Mean values for the accent measures of interest by condition, separated by accent.

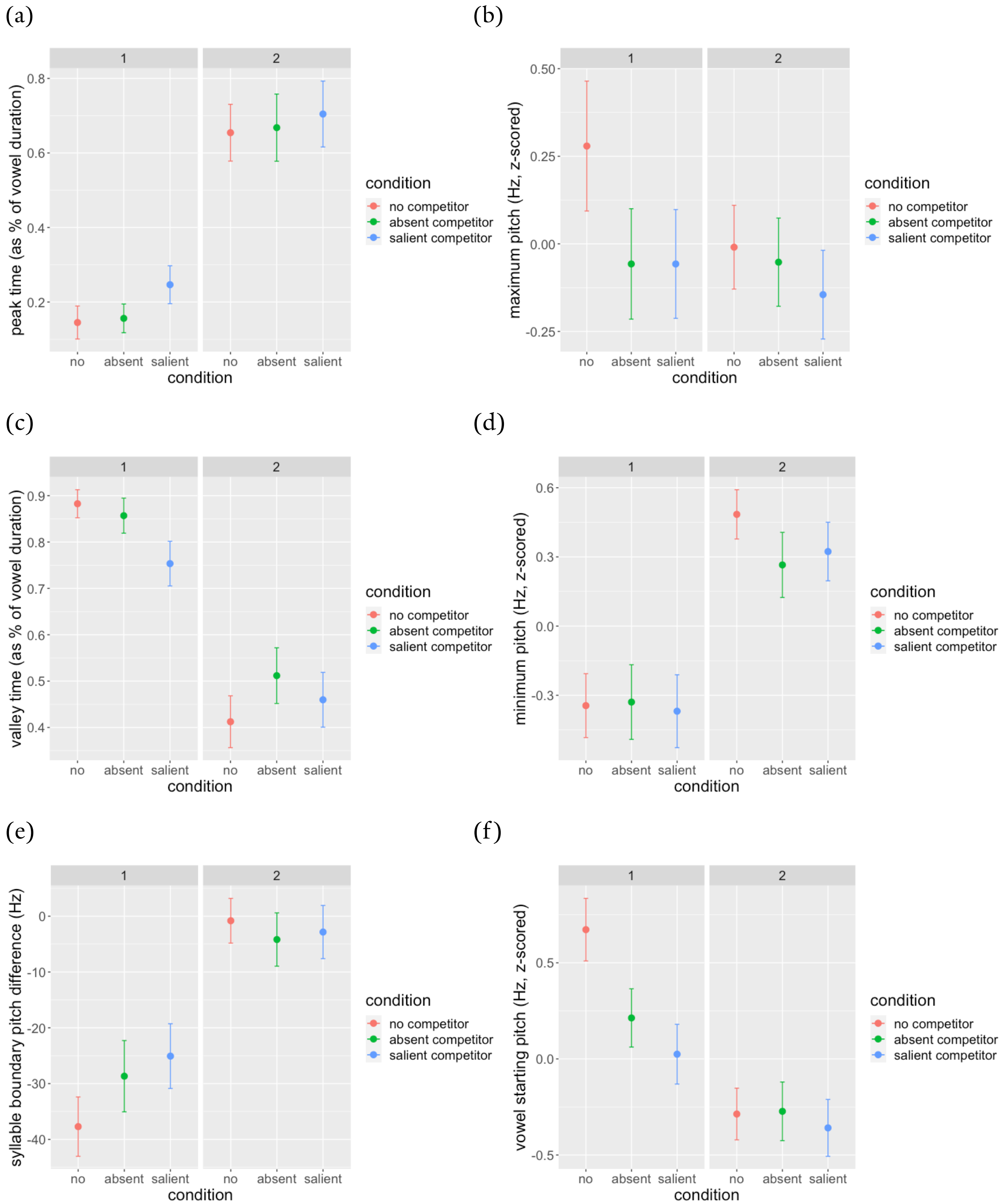
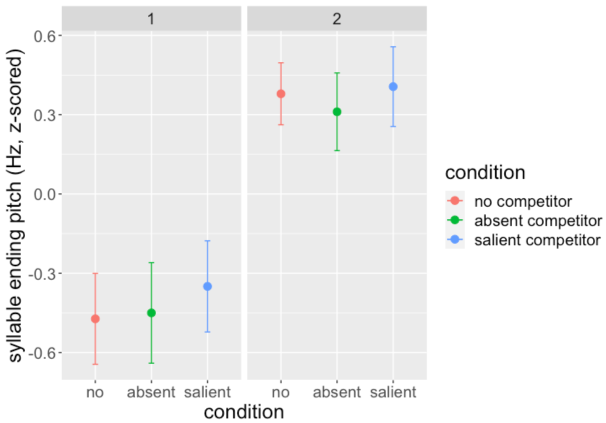

(a)
1
2
peak time (as % of vowel duration)
0.8
0.6
0.4
0.2
no absent salient no absent salient
condition
condition
no competitor
absent competitor
salient competitor
(b)
1
2
maximum pitch (Hz, z-scored)
0.50
0.25
0.00
-0.25
no absent salient no absent salient
condition
condition
no competitor
absent competitor
salient competitor
(c)
1
2
valley time (as % of vowel duration)
0.9
0.8
0.7
0.6
0.5
0.4
no absent salient no absent salient
condition
condition
no competitor
absent competitor
salient competitor
(d)
1
2
minimum pitch (Hz, z-scored)
0.6
0.3
0.0
-0.3
no absent salient no absent salient
condition
condition
no competitor
absent competitor
salient competitor
(e)
1
2
syllable boundary pitch difference (Hz)
0
-10
-20
-30
-40
no absent salient no absent salient
condition
condition
no competitor
absent competitor
salient competitor
(f)
1
2
vowel starting pitch (Hz, z-scored)
0.5
0.0
-0.5
no absent salient no absent salient
condition
condition
no competitor
absent competitor
salient competitor

Figure 7 (a-f).

(g)



(h)

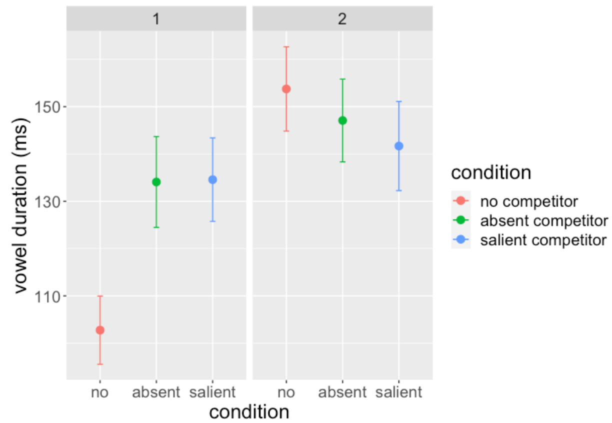


Figure 7. Means with 95% confidence intervals by condition, separated by accent, for the following measures: (a) peak time, (b) maximum pitch, (c) valley time, (d) minimum pitch, (e) syllable boundary pitch difference, (f) vowel starting pitch, (g) syllable ending pitch, (h) vowel duration.

Measure	**Accent**	**Model Estimate (Absent Comp.)**	**Model estimate (Salient Comp.)**	χ^2	**df**	***p***
peak time (% of vowel duration)	1	1.29	10.28	12.33	2	0.002 **
	2	>0.001	>0.001	1.06	2	0.59
maximum pitch (Hz)	1	-3.32	-4.60	5.19	2	0.07 .
	2	-0.24	-2.46	3.36	2	0.19
valley time (% of vowel duration)	1	-2.54	-12.88	24.57	2	>0.001***
	2	10.09	4.71	7.42	2	0.02 *
minimum pitch (Hz)	1	0.14	-1.18	0.64	2	0.73
	2	-3.22	-2.68	5.60	2	0.06 .
syllable boundary pitch difference (Hz)	1	9.02	12.33	10.90	2	0.004 **
	2	-3.37	-1.42	1.45	2	0.49
vowel starting pitch (Hz)	1	-8.41	-11.30	33.57	2	>0.001***
	2	0.07	-1.38	0.84	2	0.66
syllable ending pitch (Hz)	1	0.91	1.72	0.36	2	0.83
	2	-1.36	-0.47	0.48	2	0.79
vowel duration (ms)	1	31.02	31.68	38.52	2	>0.001***
	2	-7.11	-12.25	4.20	2	0.12

Table 6. For each accent measure, an experimental model a fixed effect of condition and a random intercept of subject was compared against a baseline model with no fixed effect. Experimental model estimates and results of experimental-baseline model comparisons for each measure are presented.

3.2.1 Peak time

For accent 1, peak time is not significantly different between the NO COMPETITOR condition and the ABSENT COMPETITOR condition, but it is significantly later in the SALIENT COMPETITOR condition relative to the NO COMPETITOR condition (estimate = 10.28% of vowel duration). Condition does not significantly affect peak time for accent 2. This indicates that along the dimension of peak time, minimal pair competition is associated with productions of accent 1 that are *more similar* to those of accent 2, but only when the competitor is contextually salient.

3.2.2 Maximum pitch

For accent 1, maximum pitch is not significantly different between the NO COMPETITOR condition and the ABSENT COMPETITOR condition; maximum pitch is, however, significantly lower in the SALIENT COMPETITOR condition relative to the NO COMPETITOR condition (estimate = 4.6 Hz). Condition does not significantly affect maximum pitch for accent 2. Along the dimension of maximum pitch, thus, minimal pair competition is associated with productions of accent 1 that are *more similar* to those of accent 2, but only when the competitor is contextually salient.

3.2.3 Valley time

For accent 1, valley time is not significantly different between the NO COMPETITOR condition and the ABSENT COMPETITOR condition, but it is significantly earlier in the SALIENT COMPETITOR condition relative to the NO COMPETITOR condition (estimate = -12.88% of vowel duration). For accent 2, valley time is significantly later in the ABSENT COMPETITOR condition relative to the NO COMPETITOR condition (estimate = 10.09% of vowel duration), but there is no significant difference in valley time between the NO COMPETITOR condition and the SALIENT COMPETITOR condition. Along this dimension, minimal pair competition is associated with bidirectional convergence between accent 1 and accent 2. For accent 1, this occurs only when the competitor is contextually salient. For accent 2, this occurs only when the competitor is *not* contextually salient.

3.2.4 Minimum pitch

Accent 1 is not significantly different in minimum pitch between conditions. For accent 2, there is a marginally significant effect of condition on minimum pitch; minimum pitch is lower in the ABSENT COMPETITOR condition relative to the NO COMPETITOR condition (estimate = -3.22 Hz) and in the SALIENT COMPETITOR condition relative to the NO COMPETITOR condition (estimate = -2.68 Hz). Minimum pitch of accent 2, thus, converges towards accent 1-like values in words with minimal pair competitors.

3.2.5 Syllable boundary pitch difference

For accent 1, syllable boundary pitch difference is significantly less negative in both the ABSENT COMPETITOR condition (estimate = 9.02 Hz) and the SALIENT COMPETITOR condition (estimate = 12.33 Hz) relative to the NO COMPETITOR condition. For accent 2, condition is not a significant predictor of syllable boundary pitch difference. This means that along this dimension, accent 1 converges towards accent 2-like values in words with minimal pair competitors and that the magnitude of this convergence effect is greater when the competitor is contextually salient.

3.2.6 Vowel starting pitch

For accent 1, vowel starting pitch is significantly lower in both the ABSENT COMPETITOR condition (estimate = -8.41 Hz) and the SALIENT COMPETITOR condition (estimate = -11.30 Hz) relative to the NO COMPETITOR condition. Vowel starting pitch is not significantly different between conditions for accent 2. Thus, vowel starting pitch for accent 1 converges towards accent 2-like values in words with minimal pair competitors. The magnitude of this convergence is greater when the competitor is contextually salient.

3.2.7 Syllable ending pitch

Syllable ending pitch is not significantly different between conditions for accent 1 or accent 2.

3.2.8 Vowel duration

For accent 1, vowel duration is significantly longer in both the ABSENT COMPETITOR condition (estimate = 31.02 ms) and the SALIENT COMPETITOR condition (estimate = 31.68 ms) than in the NO COMPETITOR condition. For accent 2, inclusion of experimental condition in the model did not significantly improve model fit. Along the dimension of vowel duration, thus, accent 1 words with minimal pair competitors are more similar to accent 2.

3.2.9 Summary of accent analysis

Accent 1 and accent 2 are differentiated along numerous dimensions of speech: peak time, maximum pitch, valley time, minimum pitch, syllable boundary pitch difference, vowel starting pitch, syllable ending pitch, and vowel duration. Along all of these dimensions except syllable ending pitch, the difference between accent 1 and accent 2 is *less* in minimal pairs than in words without minimal pair competitors. Values for accent 1 on six of these dimensions (all with the exception of minimum pitch and syllable ending pitch) are significantly closer to values for accent 2 in words with minimal pair competitors than in words without such competitors. For syllable boundary pitch difference, vowel starting pitch, and vowel duration, convergence towards accent 2 occurs across in all accent 1 words with tonal minimal pair competitors, and the magnitude of the convergence effect is greater when the competitor is contextually salient. For peak time, maximum pitch, and valley time, this convergence effect occurs only when the competitor is contextually salient. Minimal pair competition is also associated with convergence of accent 2 values towards accent 1 values, although only for valley time and minimum pitch. Minimum pitch of accent 2 is more similar to that of accent 1 across all words with minimal pair competitors. Valley time of accent 2 converges towards that of accent 1 only when the competitor is *not* salient.

Taken together, these results indicate that minimal pair competition leads to convergence, rather than divergence, between key components of the contours of accent 1 and accent 2. I discuss the significance of this finding in §4.

4 Discussion

The results of this experiment diverge from previous findings for these same phonetic dimensions in other languages. I find no effect of minimal pair competition on VOT of word-initial voiceless stops in South Swedish. This result stands in contrast with the results of prior experimental work and corpus analysis in English, which indicate a *dissimilation* (hyperarticulation) effect of minimal pair competition on VOT values that is magnified when competitors are contextually salient (Baese-Berk & Goldrick, 2009; Buz et al., 2016; Schertz, 2013; Wedel et al., 2018). Moreover, whereas tonal minimal pair competition has been found to lead to tonal divergence in Cantonese (Chow, 2020), I report that South Swedish demonstrates the opposite effect: tonal *convergence*.

Before unpacking the primary results of interest, I briefly note that I find a much larger range of average voiceless VOT values across speakers than has previously been reported for Swedish; whereas Helgason and Ringen (2008) found speaker averages for the voiceless series to fall between 54 and 70 ms, I find speaker averages ranging from 51.80 ms to 102.53 ms. Several factors could lie at the root of this divergence. It is possible that the difference in experiment mode between the two studies—in-person in Helgason and Ringen's study versus online in the present study—prevents direct comparison between VOT ranges. Alternatively, it may be that South Swedish exhibits a wider range of speaker-average VOT values than Central Swedish, the dialect investigated by Helgason and Ringen. Finally, it could be that the size of Helgason and Ringen's participant pool (six participants) was not large enough to capture the full range of VOT variation in Swedish; the present study evaluates approximately four times the number of participants.

Now, regarding the effects of minimal pair competition, there are numerous possible reasons for the discrepancies between the results of the present study and those of previous studies, which I consider in turn.

First, it could be the case that the present study has design flaws that render it incapable of detecting contrastive hyperarticulation. The most notable difference between the design of the present study and that of prior experiments utilizing the simulated partner paradigm (Buz et al., 2016; Seyfarth et al., 2016) is the inclusion of disambiguating explanations below the target word and onscreen distractors, necessitated in this study by the fact that almost all of the accent minimal pairs in Swedish are homographs. It is possible that co-activation of the words in these onscreen explanations—facilitated through visual input—interfered with planning of VOT or pitch contours.

But such interference, which would be expected to vary from trial to trial depending on the properties of the words in the explanations, cannot alone account for the systematic, context-modulated convergence effect of minimal pair competition on numerous components of the accent contours. The robust directionality of the effect suggests that the present experiment is, in fact, capable of picking up on non-random lexical-phonetic variation. With regards to the lack of significant effect on VOT, it remains possible that there is insufficient data. I note that previous online studies have reported lexical-phonetic variation associated with minimal pair competition regardless of whether the conversation partner is simulated (Buz et al., 2016; Seyfarth et al., 2016) or real (Stern & Shaw, 2023a).

Second, it might be that Swedish, unlike other languages, lacks contrastive hyperarticulation across the board. While this is theoretically possible, I consider this to be a highly unlikely scenario. Whether lexical-phonetic variation associated with minimal pair competition is driven by long-term production-perception feedback (Wedel, 2006), audience design (e.g., Buz et al., 2016), or speaker-internal patterns of activation (e.g., Baese-Berk & Goldrick, 2009), it is unclear how and why Swedish would come to differ from other languages in this way. Future work should seek to confirm this by testing non-marginal, overlapping-category contrasts in Swedish (such as vowel contrasts). Such contrasts are predicted to be hyperarticulated in minimal pairs. Furthermore, Swedish *does* exhibit effects of minimal pair competition on pitch accent contours as is observed in previous studies, albeit in the opposite direction. This weakens the case that Swedish might simply be unique.

It is much more likely that something about the phonological characteristics of these two particular contrasts in South Swedish is responsible for the observed results. I propose that the attested differences in the effect of minimal pair competition—a divergence effect for English VOT and Cantonese tone, no effect for Swedish VOT, and a convergence effect for Swedish accent contours—are consistent with differing patterns of *competitor inhibition* for each contrast of interest. In this view, lexical competition affects phonetic outputs through entirely speaker-internal processes (§1.2.3; cf. Baese-Berk & Goldrick, 2009), which are not inconsistent with communicative-functional, listener-oriented goals (Stern & Shaw, 2023a).

As summarized in §1.2.3, in a cascading activation model of lexical retrieval (Dell, 1986), activation of a target word involves activation of that word's constituent phonological categories, which in turn activate the other words in the lexicon in which those categories are found. In enacting the movement plan for a target word, then, a speaker must select that word over all of its co-active phonological neighbors. For words with phonological neighbors overlapping in only one segment (e.g., target *tall* versus neighbor *tysk*), selection is relatively "easy"; activation of neighbors is much lower than that of the target, as neighbors are only fed by phonological activation feedback from a single overlapping segment (Figure 8a). For words with minimal pair competitors overlapping in all but one segment or one feature (e.g., target *tolk* versus competitor *dolk*), selection is more difficult; activation of the competitor is very high, as it receives activation from all of the overlapping constituent categories in addition to the place of articulation feature of the first segment (Figure 8b). While there are benefits to developing a processing system in which minimal pair competitors become highly active during lexical retrieval—namely facilitating access to related speech plans (Goldrick, 2006)—co-activation is risky. Since a speaker can only enact one VOT or tonal plan at a time, competition between target words and minimal pair competitors along these dimensions must be resolved to ensure "coherent, organized behavior" (Houghton & Tipper, 1994: 53).

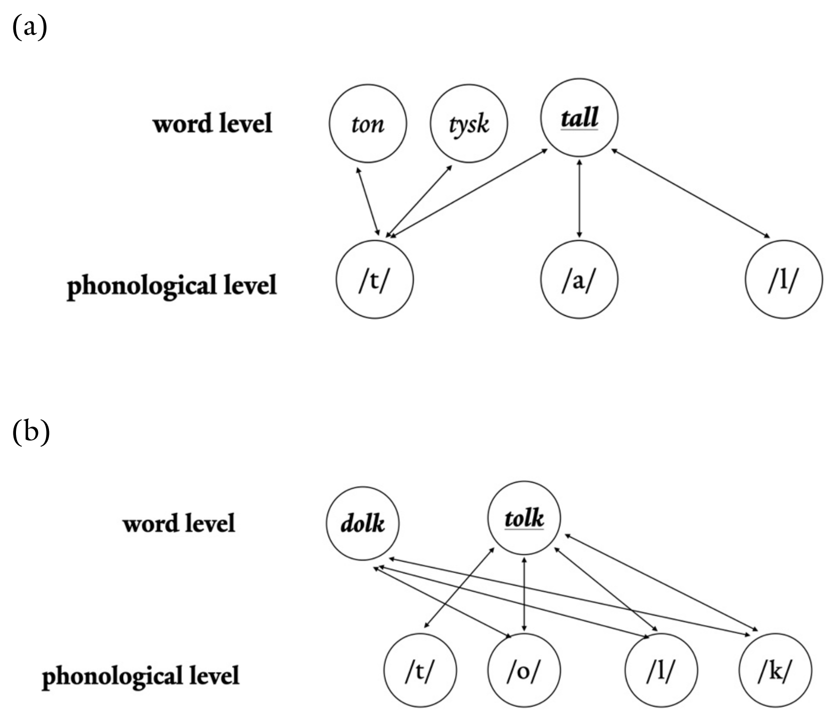


Figure 8. Cascading activation from target words (bold, underlined) without (a) and with (b) minimal pair competitors. In (a), neighbors become somewhat active, whereas in (b), competitors are highly active.

Baese-Berk and Goldrick (2009) propose that, in order to resolve lexical competition and ensure selection of the target word, the target word must receive an activation boost relative to the competitor. However, evidence of negative priming (in which access of previously ignored categories is impaired) across a range of human movements (Tipper, 1985) suggests an additional component to the neural dynamics of selection beyond *excitation* of the target word: *inhibition* of the competitor. Houghton and Tipper (1994) identify two primary evolutionary benefits of such a system. First, parallel excitation of targets and inhibition of competitors "effectively [doubles] the rate at which target and distractor (signal and noise) can be pulled apart" (Houghton & Tipper, 1994: 58) relative to an excitatory-only mechanism. Second, assuming an upper bound on activation, competitor inhibition allows for selection when both the target and competitor are highly active such that target activation cannot be further boosted. Perhaps most relevantly to the phenomenon at hand here, competitor inhibition has been demonstrated to be *selective*, such that only those components of the competitor that differ from the target are inhibited (Tipper et al., 1994).

Stern and Shaw's (2023a) model of English voicing minimal pair competition implements such a dual excitatory-inhibitory model of selection in DFT (Schöner et al., 2016). In their model, as summarized in §1.2.3, competing VOT plans for minimal pair words enter the VOT planning field as Gaussian inputs; the target (voiceless) input is excitatory and is centered at a high VOT value, and the competitor (voiced) input is inhibitory and is centered at a somewhat lower VOT value. Crucially, because English distributions for VOT are relatively close, these Gaussians overlap. This scenario, in which the overlapping lower end of the target distribution is inhibited, results in peak stabilization (i.e., a selected plan to produce VOT) at a somewhat higher VOT value than the center of the target input (Figure 9). This simulates the contrastive hyperarticulation effect observed for English VOT (Baese-Berk & Goldrick, 2009; Buz et al., 2016; Wedel et al., 2018). A similarly structured model (dual excitation-inhibition of overlapping categories) could conceivably model contrastive hyperarticulation as observed for other phonetic dimensions, such as English vowels (e.g., Wedel et al., 2018) and Cantonese tone (Chow, 2020). Importantly, inhibition in this model carries a processing cost; peak stabilization (i.e., selection time) is slower in an inhibition-excitation model of selection than in an excitation-only model (Stern & Shaw, 2023a).

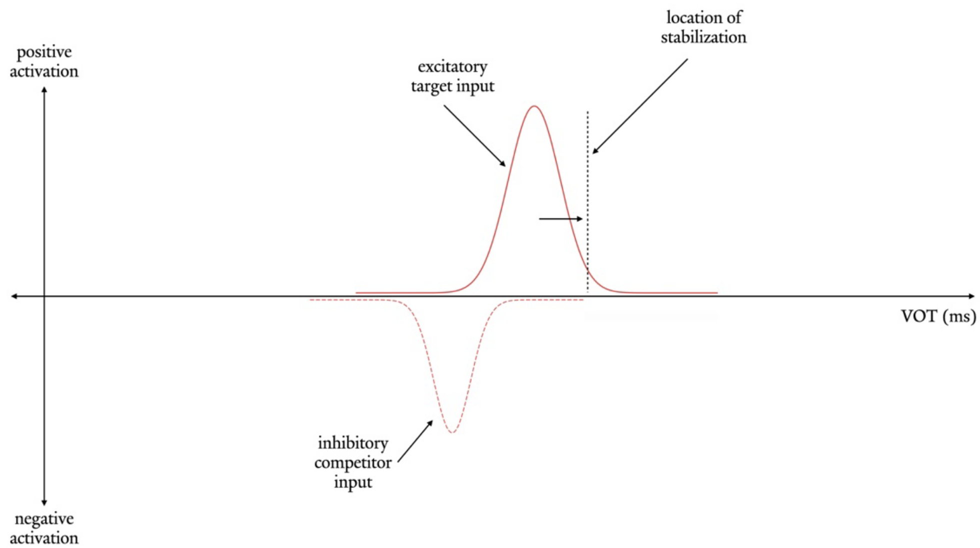


Figure 9. Possible activation dynamics for English VOT (Stern & Shaw, 2023a). Overlapping inhibitory competitor input shifts stabilization to values more extreme than the center of the target distribution.

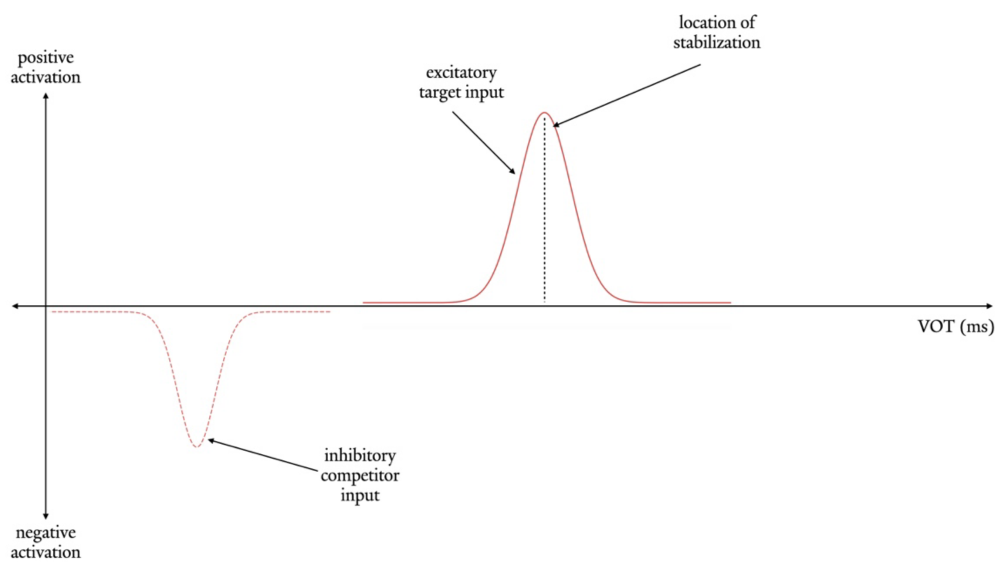


Figure 10. Possible activation dynamics for Swedish VOT. Inhibitory and excitatory inputs do not overlap, so stabilization occurs at the center of the target distribution.

Different patterns of competitor inhibition can account for the findings of the present study, namely, no effect of minimal pair competition on South Swedish VOT and convergence of pitch accent contours. As noted in §1.3.1, VOT categories are much more widely separated in Swedish—by approximately 120 ms (Beckman et al., 2011)—compared to English. In producing a voiceless target word with a minimal pair competitor, if the inhibitory competitor input does not overlap in VOT space with the excitatory target input, an activation peak is predicted to form at the center of the excitatory distribution just as if there were no minimal pair competitor (Figure 10). This is consistent with my finding that experimental condition has no effect on VOT in South Swedish.

The convergence effect of minimal pair competition on accent 1 and accent 2 contours is somewhat less straightforward. Convergence is consistent with co-activation of two *excitatory* categories, rather than an excitatory category and an inhibitory category. Neural dynamics of this type have been modeled in DFT by Stern et al. (2022) in order to capture the observation that errorful productions of voiceless stops have a somewhat lower VOT than canonical productions of voiceless stops (Alderete et al., 2021). In this model, input from a high-amplitude target distribution and input from an overlapping low-amplitude competitor distribution—both excitatory—combine to form a peak at a location intermediate between the two distribution centers (Figure 11). Applied to the South Swedish accent contrast, such neural dynamics would result in accent 1 and accent 2 contours that are more similar to each other in words with minimal pair competitors than in words without minimal pair competitors, as reported here. The role of context in modulating the convergence effect for accent 1 is consistent with increased competitor activation during trials in which the speaker views the competitor onscreen; in these trials, stronger competitor input can pull peak formation farther from the center of the target distribution than in trials in which the competitor is not contextually salient and therefore less active.

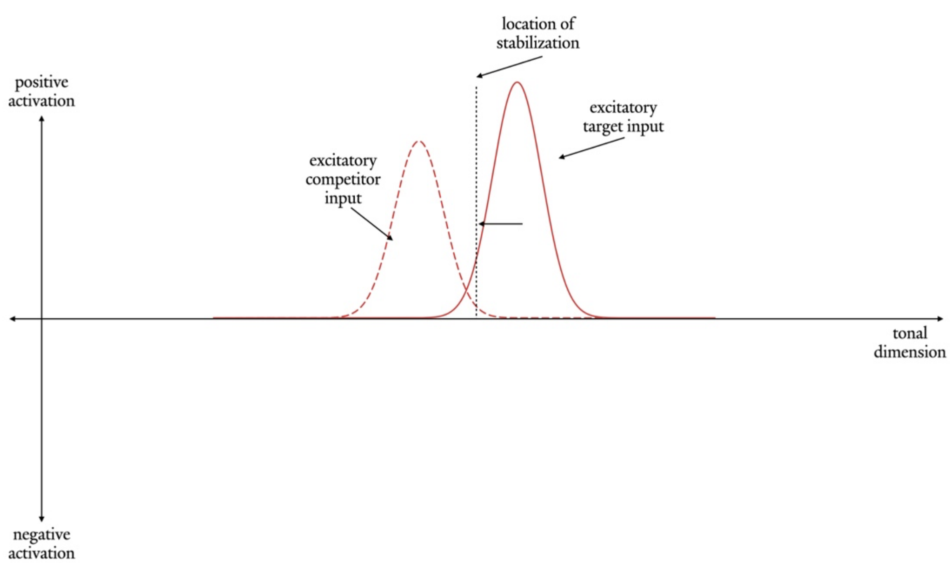


Figure 11. Possible activation dynamics for Swedish tonal dimensions. Overlapping excitatory input from a competitor shifts stabilization to a value intermediate between the centers of the two input distributions.

Why would the neural dynamics of South Swedish pitch accent production have evolved to make tonal minimal pairs more similar and therefore *more* confusable relative to words without tonal competitors? I argue that the communicative redundancy of the pitch accent contrast (§1.3.2) is crucial here. In the case of English VOT, inhibition of co-active competitors serves to mitigate the potential communicative risk of allowing selection of a target movement plan to be influenced by the competitor; more similar VOT values could lead to misperception and communicative failure. In the case of the South Swedish pitch accent contrast, however, there is little to no risk of communicative failure if competitor activation from an accent 1 competitor, for example, were to influence production of an accent 2 competitor. The vast majority of tonal minimal pairs consist of words belonging to different morphosyntactic categories, which cannot be found in the same sentential environments. Furthermore, there are very few tonal minimal pairs in total. Simply put, the risk of confusion due to pitch accent misperception is negligible. Thus, there is little listener-oriented motivation for a Swedish-acquiring speaker to develop neural dynamics that avoid convergence, especially when such inhibitory neural dynamics come with a time cost (Stern & Shaw, 2023a). There is little speaker-internal incentive, too; because the distributions of the two accents are largely predictable based on prosodic and morphosyntactic information, pitch accent selection may require less rigorous attentional demands on the speaker. Such a state of affairs obviates the need for rigorous, costly "quality control" on pitch accent output.

While it may be logical from a long-term, acquisition-informed perspective why South Swedish speakers do not inhibit competitor tonal categories in natural speech, where non-overlapping syntactic distributions render the contrast redundant, it is perhaps surprising that inhibition would also be absent in the experimental speech recorded during the present study. In the word naming task, speakers pronounce words *out of context* (i.e., without syntactic clues) with an explicit goal of disambiguating between tonal minimal pairs for a listener. Even here, when inhibition might be highly beneficial, it appears that competitor activation is excitatory. This suggests that while inhibition may be flexible according to behavioral demands (Tipper et al., 1994), such mechanisms must be learned to some extent and cannot be implemented on the fly. Or, alternatively, participants in the experiment simply did not believe in the reality of the simulated conversation partner, and their neural dynamics did not reflect those of a communicative situation.

As a final note, I briefly consider the potential relevance of the results for the debate surrounding the phonological specification of the South Swedish accent contrast (§1.3.2). Most accounts of the accent contrast in this dialect assume that the primary contrastive cue is the timing of the pitch peak; accent 1 is characterized by a pitch peak early in the first syllable, whereas accent 2 is characterized by a pitch peak late in the first syllable (Gårding & Lindblad, 1973; Riad, 2006). I find that accent 1 and accent 2 are differentiated on numerous dimensions, including not only the timing of the pitch peak, but also the maximum pitch (accent 1 > 2), the minimum pitch (accent 1 > 2), the time of the minimum pitch (accent 1 > 2), the vowel starting pitch (accent 1 > 2), the syllable ending pitch (accent 2 > 1), vowel duration (accent 2 > 1), and the syllable boundary pitch difference (a pitch fall over the syllable for accent 1 versus no fall over the syllable for accent 2). If convergence is the result of interference from a co-active, competing tonal plan, then those dimensions which show a convergence effect for accent 1 should be phonologically specified for accent 2, and vice versa. Notably, the convergence effect observed here is somewhat asymmetric; whereas accent 1 converges towards accent 2 along six of the eight measures of interest (with robust modulation by competitor salience), accent 2 converges towards accent 1 only along two measures: valley time and minimum pitch. The lack of convergence in accent 2 along other parameters, such as peak time, suggests that these parameters may not be phonologically specified for *accent 1*, since there is no evidence of a co-active competitor category interfering with accent 2. While further work is needed to confirm this hypothesis, it is possible that phonetic dimensions other than peak time—for example, amount or timing of pitch fall over the first syllable—may be better candidates for the primary (i.e., phonologized) cue(s) of contrast between the two accents.

5 Conclusion

I investigated the effects of minimal pair competition on VOT and pitch accent contours in South Swedish. I found no effect of minimal pair competition on VOT, consistent with co-activation of an excitatory target category and an inhibitory competitor category that are widely separated in feature space. I found that minimal pair competition leads to convergence of pitch accent contours along several phonetic dimensions, consistent with co-activation of target and competitor categories that are both excitatory. These results suggest that the neural dynamics of phonetic planning differ according to relevant characteristics of phonological contrasts, such as category distance and communicative redundancy. Future research should systematically explore the effects of minimal pair competition on a range of phonological contrasts with different properties and integrate these findings, including those presented here, into a unified typology of neural dynamics for phonological contrast.

Appendix

List A presents all target words from the accent trials, along with tonal minimal pair competitors when present. Superscripts denote the word accent. List B presents all target words from the VOT trials, along with voiced minimal pair competitors when present.

List A: Accent Stimuli

	Target	Gloss	Competitor	Gloss
(01)	^{1}taket	‘the roof’		
(02)	^{1}tänker	‘think (pres.)’		
(03)	^{1}bina	‘the bees’		
(04)	^{1}stämmer	‘be correct (pres.)’		
(05)	^{1}axel	‘shoulder’		
(06)	^{1}filen	‘the file’		
(07)	^{1}hamster	‘hamster’		
(08)	^{1}hästen	‘the horse’		
(09)	^{2}hundar	‘dogs’		
(10)	^{2}månen	‘the moon’		
(11)	^{2}mogen	‘mature’		
(12)	^{2}hoppar	‘jump (pres.)’		
(13)	^{2}bruten	‘broken’		
(14)	^{2}struket	‘ironed’		
(15)	^{2}ängel	‘angel’		
(16)	^{2}stöna	‘groan (inf.)’		
(17)	^{1}anden	‘the duck’	^{2}anden	‘the spirit’
(18)	^{1}backen	‘the reverse gear’	^{2}backen	‘the hill’
(19)	^{1}brister	‘burst (pres.)’	^{2}brister	‘flaws’
(20)	^{1}radar	‘radar’	^{2}radar	‘line up (pres.)’
(21)	^{1}skeden	‘the spoon’	^{2}skeden	‘shifts’
(22)	^{1}panter	‘panther’	^{2}panter	‘deposits’

(23)	^{1}slaget	‘the battle’	^{2}slaget	‘beaten’
(24)	^{1}gifter	‘marry (pres.)’	^{2}gifter	‘poisons’
(25)	^{1}världen	‘the world’	^{2}värden	‘values’
(26)	^{1}akter	‘stern (of a boat)’	^{2}akter	‘acts’
(27)	^{1}spana	‘the spas’	^{2}spana	‘observe (inf.)’
(28)	^{1}ceder	‘cedar’	^{2}seder	‘traditions’
(29)	^{1}taget	‘the grip’	^{2}taget	‘taken’
(30)	^{1}kullen	‘the litter’	^{2}kullen	‘the slope’
(31)	^{1}slutet	‘the end’	^{2}slutet	‘closed’
(32)	^{1}ljusen	‘the candles’	^{2}gjusen	‘the osprey’
(33)	^{2}tomten	‘Santa Claus’	^{1}tomten	‘the plot of land’
(34)	^{2}stället	‘the place’	^{1}stället	‘the rack’
(35)	^{2}buren	‘carried’	^{1}buren	‘the cage’
(36)	^{2}modet	‘the fashion’	^{1}modet	‘the courage’
(37)	^{2}roller	‘roles’	^{1}roller	‘roller’
(38)	^{2}knallen	‘the hill’	^{1}knallen	‘the bang’
(39)	^{2}namnen	‘the namesake’	^{1}namnen	‘the names’
(40)	^{2}leder	‘joints’	^{1}leder	‘lead (pres.)’
(41)	^{2}tanken	‘the thought’	^{1}tanken	‘the tank’
(42)	^{2}stegen	‘the ladder’	^{1}stegen	‘the steps’
(43)	^{2}bona	‘polish (inf.)’	^{1}bona	‘the nests’
(44)	^{2}hållet	‘held’	^{1}hållet	‘the direction’
(45)	^{2}regel	‘latch’	^{1}regel	‘rule’
(46)	^{2}knuten	‘tied’	^{1}knuten	‘the knot’
(47)	^{2}gripen	‘arrested’	^{1}gripen	‘the griffin’
(48)	^{2}rutten	‘rotten’	^{1}rutten	‘the route’

List B: VOT Stimuli

	Target	Gloss	Competitor	Gloss
(01)	pass	‘passport’		
(02)	päls	‘fur’		
(03)	tält	‘tent’		
(04)	tyst	‘quiet’		
(05)	kant	‘edge’		
(06)	ko	‘cow’		
(07)	kål	‘cabbage’		
(08)	kopp	‘cup’		
(09)	pil	‘willow’	bil	‘car’
(10)	par	‘couple’	bar	‘bar’
(11)	pank	‘bankrupt’	bank	‘bank’
(12)	puss	‘kiss’	buss	‘bus’
(13)	park	‘park’	bark	‘bark’
(14)	por	‘pore’	bor	‘live (pres.)’
(15)	tal	‘number’	dal	‘valley’
(16)	tagg	‘thorn’	dagg	‘dew’
(17)	tolk	‘interpreter’	dolk	‘dagger’
(18)	tass	‘paw’	dass	‘toilet’
(19)	tum	‘thumb’	dum	‘dumb’
(20)	tur	‘trip’	dur	‘major (key)’
(21)	kolv	‘piston’	golv	‘floor’
(22)	kap	‘hijinks’	gap	‘open mouth’
(23)	kul	‘fun’	gul	‘yellow’
(24)	kår	‘corps’	går	‘walk (pres.)’

References

Alderete, J., Baese-Berk, M., Leung, K., & Goldrick, M. (2021). Cascading activation in phonological planning and articulation: Evidence from spontaneous speech errors. *Cognition*, *210*, 1–5. https://doi.org/10.1016/j.cognition.2020.104577

Anwyl-Irvine, A. L., Massonié, J., Flitton, A., Kirkham, N. Z., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioural experiment builder. *Behavior Research Methods*, *52*, 388–407.

Audacity Team. (2023). *Audacity(R)* (3.4.2) [Computer software]. https://www.audacityteam.org

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, *24*(4), 527–554. https://doi.org/10.1080/01690960802299378

Bannert, R. (1998). Two thousand and one syllable in spoken Standard Swedish: Aspects of syllabification. *Reports from the Department of Phonetics, Umeå University, PHONUM*, *6*, 51–81.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Beckman, J., Helgason, P., McMurray, B., & Ringen, C. (2011). Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics*, *39*(1), 39–49. https://doi.org/10.1016/j.wocn.2010.11.001

Bell, A. (1984). Language Style as Audience Design. *Language in Society*, *13*(2), 145–204.

Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer* (6.4.06) [Computer software]. http://www.praat.org/

Bruce, G. (1970). Diphthongization in the Malmö dialect. *Working Papers/Lund University, Department of Linguistics*, *3*, 1–19.

Bruce, G. (1977). Swedish accents in sentence perspective. *Working Papers/Lund University, Department of Linguistics and Phonetics*, *12*, 61–70.

Bruce, G., & Engstrand, O. (2006). The phonetic profile of Swedish. *Language Typology and Universals*, *59*(1), 12–35. https://doi.org/10.1524/stuf.2006.59.1.12

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, *89*, 68–86. https://doi.org/10.1016/j.jml.2015.12.009

Chow, C. (2020). *Contrastive hyperarticulation of lexical tones in Cantonese* [Master's thesis]. University of Toronto.

Clopper, C. G., & Tamati, T. N. (2014). Effects of local lexical competition and regional dialect on vowel production. *The Journal of the Acoustical Society of America*, *136*(1), 1–4. https://doi.org/10.1121/1.4883478

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

Elert, C.-C. (1972). Tonality in Swedish: Rules and a List of Minimal Pairs. In E. S. Firchow (Ed.), *Studies for Einar Haugen* (pp. 151–173). De Gruyter. https://doi.org/10.1515/9783110879131-015

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, *66*(4), 789–806. https://doi.org/10.1016/j.jml.2011.11.006

Gårding, E. (1977). The Scandinavian word accents. *Working Papers/Lund University, Department of Linguistics and Phonetics*, *8*, 1–122.

Gårding, E., & Lindblad, P. (1973). Constancy and variation in Swedish word accent patterns. *Working Papers/Lund University, Department of Linguistics and Phonetics*, *7*, 36–110.

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes*, *21*(7–8), 817–855. https://doi.org/10.1080/01690960600824112

Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649–683. https://doi.org/10.1080/01690960500181332

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511616983

Gussenhoven, C., & Bruce, G. (1999). 4. Word prosody and intonation. In H. V. D. Hulst (Ed.), *Eurotyp* (Vol. 4, pp. 233–272). Mouton de Gruyter. https://doi.org/10.1515/9783110197082.1.233

Hall, K. C. (2012). Phonological Relationships: A Probabilistic Model. *McGill Working Papers in Linguistics*, *22*(1), 1–14.

Hall, K. C. (2013). A typology of intermediate phonological relationships. *The Linguistic Review*, *30*(2), 215–275. https://doi.org/10.1515/tlr-2013-0008

Haugen, E. (1967). On the Rules of Norwegian Tonality. *Language*, *43*(1), 185–202.

Helgason, P., & Ringen, C. (2008). Voicing and aspiration in Swedish stops. *Journal of Phonetics*, *36*(4), 607–628. https://doi.org/10.1016/j.wocn.2008.02.003

Houghton, G., & Tipper, S. P. (1994). A model of inhibitory mechanisms in selective attention. In D. Dagenbach & T. Carr (Eds.), *Inhibitory mechanisms in attention, memory and language* (pp. 53–113). Academic Press.

Jeong, C., & Wedel, A. (2023). The Effect of Cue-specific Lexical Competitors on Hyperarticulation of VOT and F0 Contrasts in Korean Stops. *Proceedings of the Annual Meetings on Phonology*, *10*, 1–2. https://doi.org/10.3765/amp.v10i0.5438

Kendi, A. A., & Khattab, G. (2019). Acoustic Properties of Foreigner Directed Speech. *Proceedings of the 19th International Congress of Phonetic Sciences*, 810–814.

Lahiri, A., Wetterlin, A., & Jönsson-Steiner, E. (2005). Lexical specification of tone in North Germanic. *Nordic Journal of Linguistics*, *28*(1), 61–96. https://doi.org/10.1017/S0332586505001320

Leira, V. (1998). Tonempar i bokmål. *Norskrift: Tidsskrift for Nordisk Språk Og Litteratur*, *95*, 49–86.

Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8

Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. WORD, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spoken Words: The Neighborhood Activation Model: *Ear and Hearing*, *19*(1), 1–36. https://doi.org/10.1097/00003446-199802000-00001

Marklund, E., & Gustavsson, L. (2020). The Dynamics of Vowel Hypo- and Hyperarticulation in Swedish Infant-Directed Speech to 12-Month-Olds. *Frontiers in Communication*, *5*, 1–15. https://doi.org/10.3389/fcomm.2020.523768

Munson, B., & Solomon, N. P. (2004). The Effect of Phonological Neighborhood Density on Vowel Articulation. *Journal of Speech, Language, and Hearing Research*, *47*(5), 1048–1058. https://doi.org/10.1044/1092-4388(2004/078)

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology* (pp. 101–139). Mouton de Gruyter.

R Core Team. (2024). *R: A language and environment for statistical computing* [Computer software].

Riad, T. (1998). The Origin of Scandinavian Tone Accents. *Diachronica*, *15*(1), 63–98. https://doi.org/10.1075/dia.15.1.04ria

Riad, T. (2006). Scandinavian accent typology. *Language Typology and Universals*, *59*(1), 36–55. https://doi.org/10.1524/stuf.2006.59.1.36

Riad, T. (2013). *The phonology of Swedish*. Oxford Academic.

Roll, M. (2022). The predictive function of Swedish word accents. *Frontiers in Psychology*, *13*, 1–11. https://doi.org/10.3389/fpsyg.2022.910787

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics*, *41*(3–4), 249–263. https://doi.org/10.1016/j.wocn.2013.03.007

Schöner, G., Spencer, J. P., & DFT Research Group. (2016). *Dynamic Thinking: A Primer on Dynamic Field Theory*. Oxford University Press.

Seyfarth, S., Buz, E., & Jaeger, T. F. (2016). Dynamic hyperarticulation of coda voicing contrasts. *The Journal of the Acoustical Society of America*, *139*(2), 31–37. https://doi.org/10.1121/1.4942544

Stern, M. C., Chaturvedi, M., & Shaw, J. A. (2022). A dynamic neural field model of phonetic trace effects in speech errors. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*, 3411–3417.

Stern, M. C., & Shaw, J. A. (2023a). Neural inhibition during speech planning contributes to contrastive hyperarticulation. *Journal of Memory and Language*, *132*, 1–16. https://doi.org/10.1016/j.jml.2023.104443

Tipper, S. P. (1985). The Negative Priming Effect: Inhibitory Priming by Ignored Objects. *The Quarterly Journal of Experimental Psychology Section A*, *37*(4), 571–590. https://doi.org/10.1080/14640748508400920

Tipper, S. P., Weaver, B., & Houghton, G. (1994). Behavioural Goals Determine Inhibitory Mechanisms of Selective Attention. *The Quarterly Journal of Experimental Psychology Section A*, *47*(4), 809–840. https://doi.org/10.1080/14640749408401098

Wedel, A. B. (2006). Exemplar models, evolution and language change. *The Linguistic Review*, *23*(3), 247–274. https://doi.org/10.1515/TLR.2006.010

Wedel, A. B., Nelson, N., & Sharp, R. (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language*, *100*, 61–88. https://doi.org/10.1016/j.jml.2018.01.001

Wetterlin, A. (2007). *The Lexical Specification of Norwegian Tonal Word Accents* [Doctoral dissertation]. Universität Konstanz.

Wetterlin, A. (2010). *Tonal Accents in Norwegian: Phonology, Morphology and Lexical Specification* (Vol. 535). Walter de Gruyter.

Wetterlin, A., Jönsson-Steiner, E., & Lahiri, A. (2007). Tones and loans in the history of Scandinavian. In T. Riad & C. Gussenhoven (Eds.), *Tones and Tunes: Vol. Volume 1: Typological Studies in Word and Sentence Prosody* (pp. 353–375). Mouton de Gruyter.