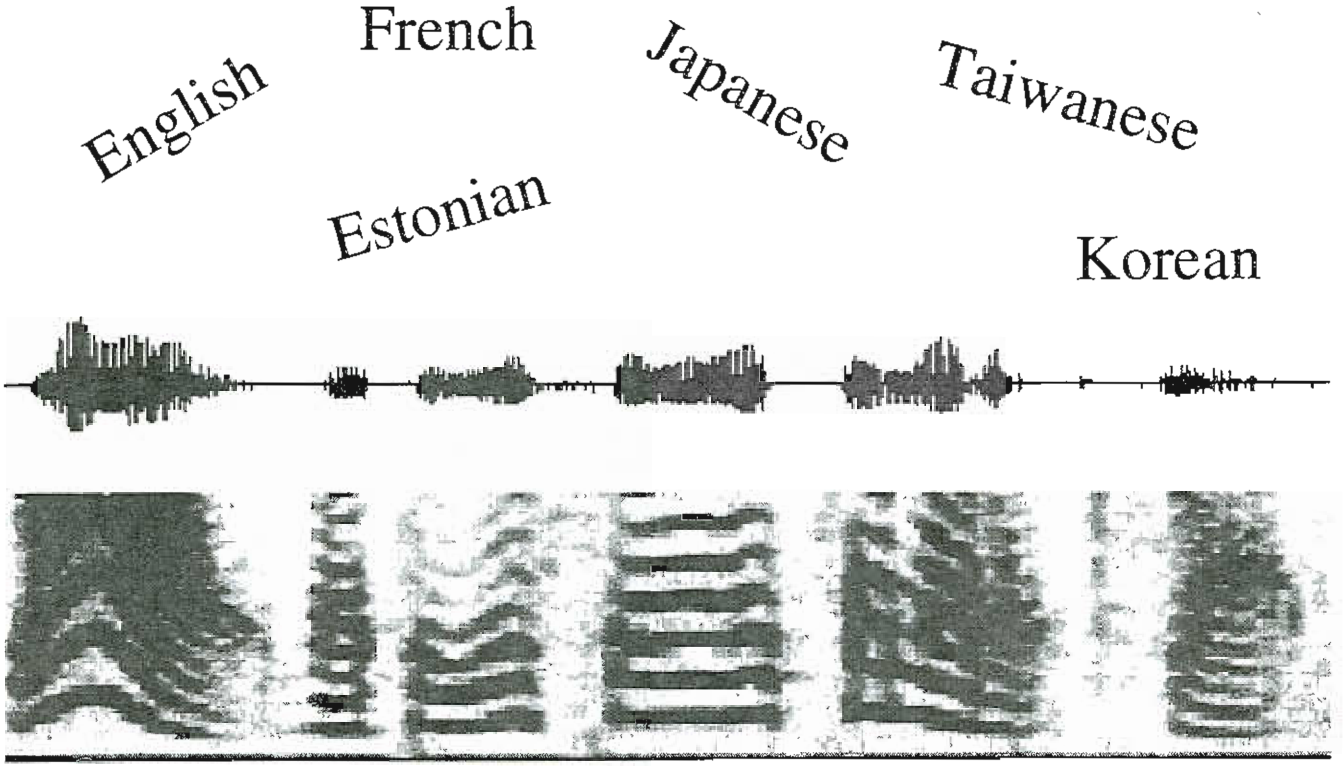
Title
WPP, No. 92

Publication Date
1996-08-01



UCLA Working Papers in Phonetics 92

August 1996

UCLA Working Papers in Phonetics

Number 92

August 1996

Table of Contents

A Phonetically-based Optimality-Theoretic Account of Consonant Reduction in Taiwanese

Chai-Shune Hsu

1. Introduction

It has often been suggested (Martinet 1952, Lindblom 1990, among others) that speech represents a compromise between two opposing goals: ease of articulation, which requires the minimization of effort during speech production, and ease of perception, which demands maximization and preservation of contrasts. In their discussion of phonetic universals in the structure of consonant systems, Lindblom and Maddieson (1988) arrive at the generalization in (1) after surveying phoneme inventories of the 317 languages in UPSID[1]:

1. "Consonant inventories tend to evolve so as to achieve maximal perceptual distinctiveness at minimum articulatory cost" (p. 72, their (6)).

In other words, small inventories will consist only of "phonetically natural" articulations. As the inventory size increases (i.e., more contrasts are implemented), the more difficult articulations will be recruited to ensure sufficient perceptual distinctiveness of the segments, but only after the phonetically natural articulations have been saturated.

Similar strategies as stated in (1)--maximizing perceptual distinctiveness at minimum articulatory cost--can be observed between subsets of a language's consonant inventories. For example, Taiwanese has the word-initial consonants in (2a) and the non-word-initial consonants in (2b) (adapted from Lin 1989: 248-249 and Zhang 1983 : 7):

2. Taiwanese Consonant Inventory
   a. Word-initial consonants:

b/m[2]	l/n	(dz)[3]	g/ŋ	
p	t	ts	k	
p^h	t^h	ts^h	k^h	
		s		h

   b. Word-final (= non-initial) consonants: (unreleased; see text for discussion)

P	T		K	ʔ
m	n		ŋ	

At a first approximation, the inventories in (2a) and (2b) certainly attest to the generalization in (1): The inventory in (2b) is about half the size of the inventory in (2a), and accordingly, the "less natural" articulations such as aspiration and voicing in obstruents are found only in the larger inventory in (2a). However, there is one important asymmetry between Taiwanese word-initial

---

[1] UPSID stands for "UCLA Phonological Segment Inventory Database." The languages in UPSID are selected "so as to approximate 'a properly constructed quota sample on a genetic basis of the world's extant languages' (Maddieson 1984: 5)" (Lindblom and Maddieson 1988: 63).

[2] The voiced oral stops and nasal stops are allophones: voiced "oral" onset stops are phonetically (weakly) prenasalized, and occur before an oral vowel; nasal onsets occur before a nasal vowel. Nasality is contrastive in vowels. For example, [i] 'he/she' vs. [ĩ] 'institution', [si] 'yes' vs. [sĩ] 'to marinate'.

[3] /dz/ has merged with /l/ in some dialects. In the dialects which retain /dz/, it occurs only before high vowels, but is still contrastive with /l/, for example, [lip] 'to erect' vs. [dzip] 'to fill' (Tung 1957: 241).

and word-final consonants the inventories in (2) fail to suggest: While the word-initial consonants are realized fairly consistently across prosodic environments in the way implied by the IPA symbols in (2a), non-word-initial consonants are articulatorily (and acoustically) more variable, hence my use of the (non-IPA) symbols /P, T, K/: Pre-pausally, /P, T, K/ are unreleased and glottalized, yielding the percept of voiceless unaspirated stops[4]. Intervocalically, they are voiced and lenited. Pre-consonantally, they are optionally (perceptually) assimilated in place (and/or manner) to the following consonant (Zhang 1983: 22-23), and the voicing is variable. Implicit in the choice of the symbols /P, T, K/ is also the desire to remain agnostic as to what the "basic" allophone should be, in view of the phoneme's articulatory variability. I also refrain from choosing a basic allophone for the word-initial voiced stops.

Interpreting this articulatory variation particular to the non-word-initial position as imprecision of the articulation, we find Kohler's statement appropriate in accounting for this asymmetry in the degree of articulatory invariance of Taiwanese word-initial and word-final consonants: Kohler (1991) states that the "word-initial position has a higher signalling value for a listener and must therefore be given a more precise articulation by a speaker.... What is not very distinctive for a listener anyway may be reduced by a speaker more easily to yield to the principle of economy of effort" (Jun 1995: 37, citing Kohler 1991: 189). In addition, it has often been suggested (for example, in Brown and McNeill 1966) that the word-initial position plays an important role in lexical access in speech production. Thus, in Taiwanese, only the perceptually less salient non-initial positions allow ease-of-articulation considerations to override the ease-of-perception principle, yielding variable articulations across prosodic environments. The smaller, ease-of-articulation-governed non-initial inventory and the larger, contrast-preservation-governed word-initial inventory of Taiwanese can thus be said to conform to Lindblom and Maddieson's generalization in (1): In non-word-initial positions, where preservation of contrasts is a relatively low priority, articulatory effort is minimized; in the word-initial position, where preservation of contrasts is a relatively high priority, more articulatory effort is invested to maximize the ease of perception.

Based on the assumption that speech production represents language-specific compromises between ease of articulation and ease of perception, this paper will provide an account of final consonant lenition in Taiwanese, focusing on supraglottal gestural reduction in /P, T, K/. The claim put forth in the present paper is that word-final reduction is a general process in Taiwanese, with the aim of effort minimization. I will argue that the intervocalic coda lenition is the consequence of a general process of gestural reduction, driven by ease of articulation considerations. The place assimilation facts in Taiwanese documented by Zhang (1983) will follow naturally from a claim advanced by Jun (1995), that gestural reduction of the word-final stop (plus compensatory lengthening of the following word-initial consonant) is the articulatory mechanism responsible for perceived place assimilation.

I will argue that the lack of (perceptual) gestural reduction of pre-pausal stops is due to final glottal fortition, which enhances the percept of stophood. Furthermore, the failure of word-initial stops to spirantize suggests the presence of a constraint forbidding the weakening of word-initial segments, a constraint presumably motivated by the perceptual considerations introduced above.

---

[4] Note that articulatorily, the percept of a voiceless unaspirated stop results from very different laryngeal (and perhaps supraglottal) configurations in onsets and pre-pausal codas: Voiceless unaspirated onset stops are produced with an open glottis during oral closure (voicelessness), and upon release of the stop closure, the glottis has closed and vocal fold vibration commences (no aspiration). Utterance-final stops, on the other hand, achieve the percept of voicelessness and absence of aspiration due to glottalization, which prevents vocal fold vibration, the articulatory correlate of [voice]; glottalization during the consonantal release also prevents an audible burst, hence the lack of aspiration.

In Section 2, I will briefly outline some assumptions about Taiwanese phonology adopted in this paper, and summarize the facts I propose to account for in the present study. Section 3 reviews the parts of Jun (1995) relevant to the current study. Section 4 describes a series of oral pressure experiments which provide evidence of pre-consonantal lenition in Taiwanese. Section 5 proposes an account of lenition in Taiwanese, within the framework of Phonetically-Based Optimality Theory.

2. Some aspects of Taiwanese phonology

2.1 *Syllables*

Taiwanese syllables allow maximally one consonantal onset and one consonantal coda, but neither onsets nor codas are obligatory. This means that all consonant sequences CC are necessarily heterosyllabic, with C1 being the coda and C2 being the onset. Furthermore, in a VCV sequence, the intervocalic consonant may be an onset (V.CV) or a coda (VC.V); the latter syllabification is the result of vowel-initial suffixation.

As is the case of most Chinese languages, each word in Taiwanese is a syllable, and each syllable is a word, with the exception of the suffixes /-a/[5] and /-e/[6] which are bound morphemes assumed here to be onsetless syllables.

2.2 *The realization of Taiwanese word-initial and word-final stops*

This section will summarize the observations about Taiwanese this paper proposes to account for, with the focus on the supraglottal articulations of consonants. Although the discussion mentions only the codas /P, T, K/, similar assumptions about the supralaryngeal stricture can be made about the nasal codas /m, n, ŋ/.

Intervocalic position:

In Taiwanese, word-final /P, T, K/ lenite in intervocalic position, with the reduced /T/ most often perceived as a flap/tap, and the others as continuants:

3. aP a[7] -> aβ̞a 'little box'
kʰuT a -> kʰuɾ̞a 'little hole'
tiK a -> tiɣ̞a 'bamboo'

The degree of intervocalic lenition is variable (as my use of the lowering diacritic [̞] suggests), and is always accompanied by voicing. Moreover, while word-final lenition consistently occurs before vowel-initial suffixes (/-e/, /-a/), it is more variable before vowel-initial words, which sometimes acquire an initial glottal stop /ʔ/ in slower, more careful speech, accompanied by glottalization of the preceding (unreleased) word-final stop. In (4), the word /aP/[8] stands for 'box':

---

[5] The suffix /-a/ has several meanings. For example, it may be a diminutive marker, as in /ap a/ 'little box (*lit.*, box A)'; it may be suffixed to familiar objects, as in /kam a/ 'oranges (*lit.*, orange A)'; it may also mark a pejorative tone, as in /kiŋ tsat a/ 'the (little) cops (*lit.*, police A)'.

[6] Like /-a/, the suffix /-e/ also has many usages. It can be used as a general counter/classifier, as in /tsit e laŋ/ 'one person (*lit.*, one E person)'; it is an agentive marker, as in /li e tsʰeʔ/ 'your book (*lit.*, you E book)'; it may be used as a relativiser, as in /li koŋ e ue/ 'what you have said (*lit.*, you say E speech)'; it is also a nominalizer, as in /aŋ sik e/ 'the red one (*lit.*, red color E)'.

[7] Tones will not be indicated in this paper, as they are irrelevant for the present study.

[8] The word for box is /aʔ/ in some dialects. This variation will not be relevant in the present discussion.

4.	tsapˀ ʔaP	~	tsaβaP	'ten boxes'
	tsitˀ ʔaP	~	tsiɾaP	'one box'
	lakˀ ʔaP	~	laɣaP	'six boxes'

Intervocalic lenition is particular to word-final position. Word-initial stops never show the alternation between non-continuants and continuants in the intervocalic environment. Focusing on the intervocalic word-initial consonants below, (5) and (6) show that voiceless stops undergo neither spirantization nor voicing in the intervocalic environment; (7) shows that voiced stops also do not undergo spirantization in the same environment:

5.	pe paʔ	[pe paʔ]	'eight hundred'	*[pe ɸaʔ]	*[pe βaʔ]
	pe tã	[pe tã]	'eight stands'	*[pe ɾ̥ã]	*[pe ɾã]
	pe kaʔ	[pe kaʔ]	'eight hectare'	*[pe χaʔ]	*[pe ɣaʔ]
6.	pe pʰa	[pe pʰa]	'eight lights'	*[pe ɸa]	*[pe βa]
	pe tʰaʔ	[pe tʰaʔ]	'eight stacks'	*[pe ɾ̥aʔ]	*[pe ɾaʔ]
	pe kʰa	[pe kʰa]	'eight players'	*[pe χa]	*[pe ɣa]
7.	pe ba	[pe ba]	'eight yards'	*[pe βa]	
	li li	[li li]	'twenty-two'	*[li ɾi]	
	li gɔ	[li gɔ]	'twenty-five'	*[li ɣɔ]	

Place assimilation:

Word-final stops optionally assimilate in place to the following consonant, while word-initial stops never undergo progressive place assimilation:

8.	/tsaP pɔ/	[tsappɔ]	'ten steps'	
	/tsaP te/	[tsapte] ~ [tsatte]	'ten pieces'	*[tsappe]
	/tsaP kʰɔ/	[tsapkʰɔ] ~ [tsakkʰɔ]	'ten dollars'	*[tsappʰɔ]
	/tsʰiT pɔ/	[tsʰitpɔ] ~ [tsʰippɔ]	'seven steps'	*[tsʰittɔ]
	/tsʰiT te/	[tsʰitte]	'seven pieces'	
	/tsʰiT kʰɔ/	[tsʰitkʰɔ] ~ [tsʰikkʰɔ]	'seven dollars'	*[tsʰittʰɔ]
	/laK pɔ/	[lakpɔ] ~ [lappɔ]	'six steps'	*[lakkɔ]
	/laK te/	[lakte] ~ [latte]	'six pieces'	*[lakke]
	/laK kʰɔ/	[lakkʰɔ]	'six dollars'	

Place assimilation is not limited to consonants preceding a stop; it occurs also before nasals and sibilants: /laK mĩ/ 'six tokens' [lakmĩ] ~ [lammĩ]; /laK sɤ/ 'Six-Keys' [laksɤ] ~ [lassɤ][9].

It is my impression that coronal codas are the most likely to undergo place assimilation, while labial codas are the least likely to undergo place assimilation. This observation is supported by the frequency of the documented assimilated codas in Zhang (1983: 22-23).

Pre-pausally:

Pre-pausal codas undergo glottalization, and lenition is not (perceptually) attested in this environment. Here, pre-pausal position includes utterance-final position and utterance-medial but pre-pausal positions in careful speech.

---

[9] The continuancy feature of C1 is also assimilated by C2 in this example.

Word-initially:

Finally, I would like to re-iterate that word-initial stops never undergo lenition; supraglottal closure always obtains whether the stop occurs in phrase-initial or phrase-medial positions.

Consistency in intervocalic lenition:

Note from the previous discussion of word-final consonants that lenition is consistent in the pre-vocalic position, but more variable in the pre-consonantal position.

In sum, word-final reduction appears to be a rather general phenomenon in Taiwanese, occurring (at least) in all phrase-/utterance-medial positions. I suggest that Jun's (1995) Optimality Theoretic analysis of place assimilation can be extended to account for reduction in Taiwanese. The next section will review the parts of Jun (1995) relevant to the analysis of Taiwanese.

3. Jun (1995)

This section will provide a partial summary of Jun (1995), a reduction-based account of place assimilation. Both articulatory and perceptual factors are considered in his study, and the opposing demands of articulation and perception are reconciled within the framework of Optimality Theory.

3.1 *Articulatory correlate of place assimilation*

Like Jun, we will focus here only on the optional, casual-speech gradient regressive place assimilation in coda-onset clusters. This place assimilation can be represented as the perceptual deletion of the target place (C1) and the compensatory lengthening of the trigger place (C2):

9. Regressive Place Assimilation: C1C2 -> C2C2

It has been suggested by Browman and Goldstein (1986; 1989; 1990; 1992; cited in Jun 1995: 29) within the framework of Articulatory Phonology that casual-speech place assimilation may be due to either temporal overlap of the coda gesture by the onset gesture, or the spatial-temporal reduction of the coda gesture. In order to isolate the particular contributions of these two articulatory mechanisms--overlap and reduction--in the perceptual loss of C1 place cues in place assimilation, Jun (1995) conducted a series of production and perception experiments on English and Korean, which differ in the range of place assimilations observed.

Adopting the methodologies and interpretations developed in Silverman and Jun (1994), Jun (1995) monitored the pressure build-up behind the lips[10] during the articulation of (Korean and English) utterances containing /pt/ and /pk/ clusters. He found that while the labial gesture [p] never reduces before [t], labial reduction sometimes obtains before [k] in Korean, but not in English. The results so far parallel the observation that /pt/ never undergoes place assimilation in both English and Korean, and that /pk/ may optionally undergo place assimilation in Korean only, in which case the cluster is perceived as [kk].

Follow-up perceptual experiments showed that whether listeners heard /p/ as assimilated in /pk/ clusters depended on whether the /p/ was reduced: Only in case of a reduced /p/ did the place cues of the labial coda remain undetected in /pk/. That is, place assimilation in /pk/ obtains if and only if the labial coda is reduced; otherwise, [pk] is perceived, irrespective of the degree of p-k overlap. Jun (1995) concludes that gestural reduction, and not overlap, plays the decisive role in perceptual assimilation.

---

[10] A more thorough discussion of the inferences of the oral pressure readings will be provided in Section 4.

Finally, he concludes with Barry (1992) and Nolan (1992) that gestural reduction in casual speech place assimilation is speaker-controlled, that is, gestural reduction in gradient place assimilation does not arise as the consequence of vocal tract constraints (p. 135).

3.2 *Typological generalizations*

Based on a more extensive typological survey of place assimilation, Jun (1995) corroborates Mohanan's (1993) cross-linguistic generalization[11] that if non-coronals undergo place assimilation, so do coronals. Moreover, based on the place assimilation patterns in Korean and the assimilation patterns in the Inuktitut dialects surveyed in Dorais (1986), he elaborates on Mohanan's generalization by suggesting that if velars are targets of place assimilation, so are labials, and if labials are targets of place assimilation, so are coronals (p. 92), i.e., coronals are more likely to undergo place assimilation than labials, which are in turn more likely to undergo place assimilation than velars. Note that in Taiwanese, although /P, T, K/ may all undergo place assimilation, and thus conform to Jun's (1995) implicational statements, labials are *less* likely than velars to be targets in assimilation, which appears to be a counter-example to the statement that velar targets implies labial targets, i.e., velars are less likely targets than labials in place assimilation.

Jun's generalization is based on the observation that in Korean, both coronal and labial codas may be targeted in place assimilation, while velars never undergo place assimilation. Furthermore, many of the languages surveyed only allow coronal codas[12] to undergo place assimilation (e.g., Catalan, English, German, Toba Batak, Yakut (p.82)). While there are languages which allow all (unreleased) codas (/p, t, k/) to be targets in place assimilation (e.g., Diola Fogny, Japanese, Malay, Nchufie, Yoruba (p.82)), no language targets non-coronals in place assimilation to the exclusion of coronals. Similarly, no language targets velars in place assimilation to the exclusion of labials (and coronals).

3.3 *Analysis*

3.3.1 *Theoretical assumptions*

In explaining the varying degree of likeliness of place assimilation targets, Jun (1995) assumes the Production Hypothesis suggested by Steriade (1993) and Byrd (1994):

10. Production Hypothesis (Jun 1995: 35, his (35)):
    Speakers make more effort to preserve the articulation of speech sounds with powerful acoustic cues, whereas they relax in the articulation of sounds with weak cues.

That is, the most favored targets in place assimilation are such because they have the weakest acoustic cues even in their non-reduced forms. Jun argues that the relative likelihood of coda targets reflects the relative robustness of their perceptual cues in a consonant cluster. In a stop (nasal or oral) consonant cluster <u>CC</u>, <u>C1</u> is usually unreleased (i.e., the release or stop burst of <u>C1</u> is inaudible) due to overlap by <u>C2</u>. In the case of an unreleased <u>C1</u>, identification of its place relies solely on the V-to-C transition, which is significantly less salient than the acoustic cues present in C-to-V transitions. Moreover, the place cues of <u>C1</u> in V-to-C transitions may be influenced by <u>C2</u> (p. 153, citing Byrd 1992, Zsiga 1992), especially upon considerable overlap by a slow <u>C2</u> gesture (e.g., non-coronal <u>C2</u>).

---

[11] Mohanan's (1993) generalizations are based on attested place assimilation patterns in English, Korean, Hindi, and Malayalam (Jun 1995: 40).

[12] Jun's (1995) generalizations include both oral and nasal codas.

Assuming that unreleased coronals have weaker place cues than unreleased non-coronals, and further, unreleased labials have weaker place cues than unreleased velars (See Jun 1995, Section 4.2.1.2: 149-151 for a detailed discussion), Jun proposes that the coronal gesture is the least likely to be preserved under unrelease; labials are the next least likely candidate of preservation; and finally, unreleased velar codas are the most likely to be preserved as a stop. Moreover, he assumes this tendency to be universal. Below, I will provide his formalization of this tendency after briefly reviewing the formal framework of Optimality Theory which both Jun (1995) and the present paper employ.

In Optimality Theory (Prince and Smolensky 1993, McCarthy and Prince 1993), the phonology is composed of a set of universal constraints. Unlike in other constraint-based theories, these constraints are violable, and they are ranked. The constraint ranking is language-specific. Thus, cross-linguistic variations result from variations in constraint ranking, and not variations in the constraint composition: The constraint ranking determines the possible outputs of a grammar. For each given input form, an infinite number of outputs are generated. In determining the winning form, these outputs are evaluated against the strictly-ranked constraint hierarchy. An output which violates a higher-ranked constraint is always disfavored to one which violates a lower-ranked constraint. (See Prince and Smolensky 1993 for a more comprehensive discussion of the procedure of output evaluation.)

Using Jun's (1995) constraints as an illustration, assume the following subset of constraints in regressive assimilation. I will only focus here on the reduction of C1 in a CC cluster, and assume that highly ranked constraints ensure the compensatory lengthening by C2 as C1 reduces, as well as the absence of progressive place assimilation (i.e., reduction of C2 plus compensatory lengthening by C1):

11. Reduction Constraint (p. 143, his Chapter 2, (2)):
    RED: Conserve articulatory effort.

12. Preservation Constraint (generalizing his Chapter 4, (6), p. 151):
    Pres(pl(X˺)): Preserve the place cues of an unreleased articulation X˺,
        where X ranges among the places dorsal, labial, and coronal.

The reduction constraint RED requires that a consonantal gesture be reduced, or lenited[13], while the preservation constraint Pres(pl(X˺)) requires that the cues for the place of articulation X be preserved; for an unreleased pre-consonantal coda, this means the articulatory gesture giving rise to the place cues for X must be unreduced: These two constraints clearly encode opposing demands. Now assume two hypothetical languages L1 and L2 which contrast in the relative ranking of these two constraints. In L1, RED outranks Pres(pl(X˺)), i.e., reduction is favored. In L2, Pres(pl(X˺)) is ranked higher than RED, i.e., preservation is favored. These two partial grammars are represented in (13), where the notation ">>" reads "outranks":

13 a. L1: RED >> Pres(pl(X˺))
   b. L2: Pres(pl(X˺)) >> RED

The tableaux in (14) and (15) evaluate the relevant competing output forms of L1 and L2. In the tableaux below, the input form is indicated on the top left cell, and the relevant candidates are indicated in the left-most column, where C1C2 represents an unreduced C1, and C2C2 represents

---

[13] Gestural deletion is an extreme form of reduction. In Jun (1995), deletion is the result of a separate, highly-ranked constraint DEL (p. 144), which requires maximal conservation of articulatory effort. Thus, I follow Jun in assuming that while the constraint RED demands gestural reduction, it cannot force gestural deletion.

a reduced C1[14]. The notation "→" is placed next to the winning candidate. The top row lists the constraints, with the convention that the constraint to the left outranks the constraint to the right. Each "*" in the cell records one violation of the constraint indicated at the top of the column for the corresponding candidate output. The notation "!" is placed before the crucial violation which leads to the candidate's demise. We are now ready to examine the tableaux below:

14. L1 (Place Assimilation)

/VC1C2V/	RED	Pres(Pl(X˺))
VC1C2V	!*	
→ VC2C2V		*

15. L2 (No Place Assimilation)

/VC1C2V/	Pres(Pl(X˺))	RED
→ VC1C2V		*
VC2C2V	!*	

Before we examine the two grammars L1 and L2 separately, note first that given the same input /VC1C2V/, the outputs commit the same number of violations of the same constraints in both languages: Focusing on the realization of the coda (C1), the output VC1C2V commits one violation of RED by failing to reduce[15], but satisfies Pres(Pl(X˺)) in doing so. In contrast, the output VC2C2V satisfies RED by reducing C1, but violates Pres(Pl(X˺)) in doing so.

Now turning our attention to the language-internal rankings in (14) and (15), we note that the constraint ranking in L1 means that the output VC1C2V violates a more highly-ranked constraint than does VC2C2V; hence, the latter candidate is the optimal one. In contrast, (15) shows that the constraint ranking in L2 means the candidate VC1C2V violates a more lowly-ranked constraint than VC2C2V, thus emerging as the optimal candidate. Optimality Theory thus provides an explicit framework within which the language-specific interaction of the conflicting demands of production and perception can be formalized.

3.3.2 *A formal account of the place assimilation typology*

Returning to the formalization of the place assimilation typology in Jun (1995), two families of constraints are relevant for determining the realization of targets in place assimilation. The first family of constraints is driven by minimization of effort considerations, embodied in the constraint RED repeated in (16). Kirchner's (1995) constraint (family) **LAZY** similarly encodes a preference for effort minimization.

16. RED: Conserve articulatory effort

The second family of constraints is driven by perceptual considerations. He proposes a universal ranking based on the acoustic salience of the place cues, assuming the Production Hypothesis. The ranking is provided in (17), in which the set of individual constraints is an instantiation of my generalized constraint in (12):

17. Universal ranking for target places (p.151, his Chapter 4, (6)):
    Pres(pl(dor˺)) >> Pres(pl(lab˺)) >> Pres(pl(cor˺))

---

[14] Recall that we are only concerned with the reduction of C1, and neither the product of progressive assimilation (C1C1) nor the result of C1 reduction with no compensatory lengthening of C2 (ØC2) are deemed possible (better) outputs.

[15] Notice that the onset C2 also commits one violation of reduction which I do not indicate here.

The universal constraint hierarchy in (17) indicates that unreleased velars are more likely to be preserved than unreleased labials and unreleased coronals. Moreover, unreleased labials are more likely to be preserved than unreleased coronals, where preservation of place cues may be interpreted to mean resistance to reduction.

Using the universal constraints and constraint hierarchy in (16) and (17), Jun (1995) accounts for the attested cross-linguistic patterns of place assimilation. Again, I focus here only on the gestural reduction of C1 in the cluster C1C2. The interaction of (16) and (17) predicts four possible patterns:

Pattern 1: **RED** >> Pres(pl(dor˺)) >> Pres(pl(lab˺)) >> Pres(pl(cor˺))

This ranking predicts that all codas (velars, labials, coronals) may be targeted in place assimilation, since RED outranks the entire range of preservation constraints which deter reduction. This pattern is attested in Diola Fogny, Japanese, Malay, Nchufie, and Yoruba. Taiwanese also seems to fall under this ranking. However, I will argue in Section 5 that the hierarchy as stated here only captures part of the place assimilation typology facts.

Pattern 2: Pres(pl(dor˺)) >> **RED** >> Pres(pl(lab˺)) >> Pres(pl(cor˺))

This ranking predicts that labial and coronal codas may be targeted in place assimilation, but velar codas never will, since its preservation constraint outranks the reduction constraint. This pattern is unambiguously attested in Korean. See Jun (1995), Section 4.3.3 (pp. 171-177) for a detailed discussion of Korean place assimilation.

Pattern 3: Pres(pl(dor˺)) >> Pres(pl(lab˺)) >> **RED** >> Pres(pl(cor˺))

This ranking predicts that only coronal codas will undergo place assimilation, since the reduction constraint is outranked by the constraints preserving the place cues of unreleased velars and unreleased labials. This pattern is unambiguously attested in Catalan, English, German, Toba Batak, and Yakut. See Jun (1995), Section 4.3.2 (pp. 166-170) for a detailed discussion of place assimilation in English.

Pattern 4: Pres(pl(dor˺)) >> Pres(pl(lab˺)) >> Pres(pl(cor˺)) >> **RED**

Finally, this ranking predicts that place assimilation will never occur, since the reduction constraint is outranked by the entire range of preservation constraints. Although Jun (1995) does not discuss this case, we can tentatively assume that all languages which disallow place assimilation of unreleased codas share this constraint hierarchy[16].

In conclusion, interaction of the reduction constraint in (16) with the universal preservation constraint hierarchy in (17) generates all and only the attested place assimilation typology surveyed by Jun (1995).

With these background assumptions in mind, we now return to the case of Taiwanese. We propose that the observations about intervocalic lenition and place assimilation can receive a unified Junnian account of gestural reduction. However, before we embark on an Optimality Theoretic analysis of Taiwanese, we must first confirm our hypothesis that coda reduction indeed occurs pre-consonantally. Section 4 describes an experiment which establishes that lenition obtains in the pre-consonantal position in Taiwanese. Section 5 provides a formal account of word-final reduction in Taiwanese.

---

[16] However, we conjecture that a close examination of cross-linguistic casual-speech data will reveal that such a language does not exist. That is, gestural reduction of pre-consonantal codas is expected to occur cross-linguistically in fast speech.

4. Experiment: Aerodynamic evidence of pre-consonantal reduction in Taiwanese

4.1 *Objective of the experiment*

Regressive place assimilation has been documented by Zhang (1983: 22-23) for Taiwanese and by Tung (1975: 239) for Xiamen, a closely related dialect. I have also observed optional casual-speech place assimilation in the speech of a number of Taiwanese speakers. Drawing from the conclusions of Jun (1995), we predict that the percept of place assimilation in casual speech is an indication of gestural reduction of the coda plus compensatory lengthening by the following (unreduced) onset; place assimilation is therefore related to intervocalic reduction. The goal of this experiment is to determine whether coda reduction indeed obtains in Taiwanese[17]. An affirmative answer to this question would suggest that word-final lenition is part of a more general phenomenon in Taiwanese, i.e., lenition is not limited to the intervocalic environment.

4.2 Hypothesis

In Taiwanese, pre-consonantal codas may be lenited in casual speech. This hypothesis is suggested by the percept of place assimilation and Jun's (1995: 139) conclusion that "gestural reduction plays the decisive role in the perceptual loss of the target in [casual-speech] place assimilation[18]."

4.3 *Methodological assumptions*

The basis of this experiment is the observation that air flowing via the glottis into the oral cavity will result in pressure build-up in the cavity if the incoming air cannot be (sufficiently) vented, as in the case of a supra-laryngeal stop closure. Section 4.3.1 discusses some expected oral pressure patterns, considering factors such as the presence or absence of incoming flow, the presence or absence of supra-laryngeal stop closure(s), and the consequences of oral cavity expansion and compression. Section 4.3.2 summarizes the methods and interpretations developed in Silverman & Jun (1994), which the present experiment adopts.

4.3.1 *Relevant factors in the interpretation of oral pressure data*

The figures below provide schematic diagrams of the oral cavity. In Figs. 1 and 2, Points A and B correspond to different places of constriction: the lips and the glottis, respectively. $\underline{P_A}$ denotes the pressure behind constriction point A, or the pressure inside the cavity enclosed by constriction points A and B. In Figs. 1a and 1b, the glottis is open, and air flows via the glottis into the oral cavity. When the lips are open, as schematized in Fig. 1a, no pressure will accumulate in the oral cavity, and the pressure behind constriction point A ($\underline{P_A}$) approximates the atmospheric pressure (to the left of point A), conventionally zero. If the lips are closed, as in Fig. 1b, the incoming air flow will result in a pressure build-up behind constriction point A: I will represent the positive increase in pressure as $\underline{+p}$.

---

[17] This experiment does not investigate the correlation between the presence/absence of gestural reduction and the perception of place assimilation, as we are only focusing here on the former variable, and not in a correlation of the two.

[18] Note that reduction is limited to the pre-consonantal position in Korean, but occurs both pre-consonantally and pre-vocalically in Taiwanese. This difference will be addressed in Section 5.

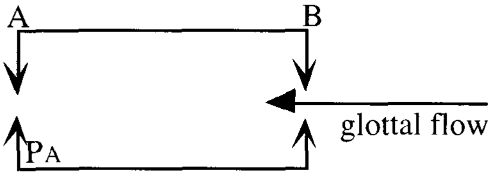


Figure 1a: glottis open; lips open

$\underline{P_A} = 0$

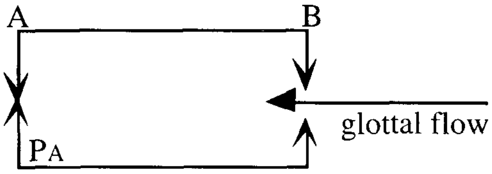


Figure 1b: glottis open; lips closed

$\underline{P_A} = +p$

In Figs. 2a and 2b, the glottis is closed, for example, during the glottal stop [ʔ] or the glottalized consonant [ˀp]. Assume the stop constriction is first formed at point B, and that the size of the cavity between A and B remains constant (for example, neither active expansion by larynx lowering nor active compression by larynx raising). In both cases, the pressure behind constriction point A is essentially zero, even when a stop closure is formed at A; this is because the glottal closure reduces glottal flow to zero, thereby removing the source of pressure build-up in the oral cavity.

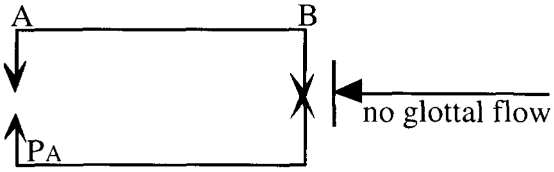


Figure 2a: glottis closed, lips open

$\underline{P_A} = 0$

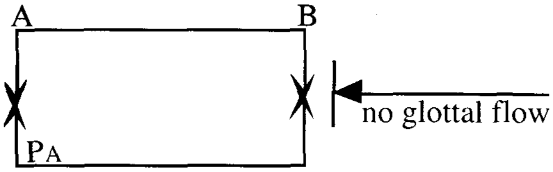


Figure 2b: glottis closed, lips closed

$\underline{P_A} = 0$

Next, consider the cases of articulatory overlap, where two supralaryngeal constrictions are formed, such as during an overlapped pk cluster. Only the cases with open glottis will be considered, since only these will be relevant for the present discussion. In the remainder of this section, the constrictions at A and B refer to the labial and velar places of articulation, respectively, while C refers to the glottis.

Fig. 3 represents an oral cavity with two sub-cavities, one between points A and B, and one between points B and C. In Fig. 3, all three places of constriction are open (i.e., neither labial, nor velar, nor glottal closure): Air flows into the oral cavity and through points B and A (relatively) freely; no pressure builds up behind either point A or point B.

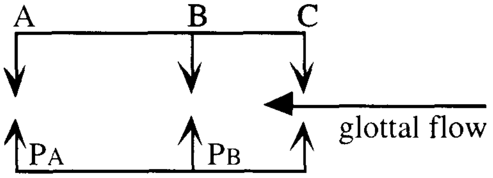


Figure 3: no labial closure, no velar closure; open glottis
$\underline{P_A} = 0$, $\underline{P_B} = 0$

Figs. 4a and 4b schematize the situations with one supralaryngeal closure (during articulatory overlap of a stop plus a lenited stop), alternating the stop closure between points A and B. In Fig. 4a, labial closure is achieved, but not velar closure. This is similar to Fig. 1b. If pressure measurements were taken behind points A and B, both should detect positive pressure, since both $\underline{P_A}$ and $\underline{P_B}$ represent the pressure in the cavity enclosed by points A and C. In contrast, the two sub-cavities in Fig. 4b have different oral pressure types. In Fig. 4b, velar closure is achieved, but not labial closure. In this case, the cavity between A and B will have zero pressure, as in Fig. 2a, since there is no air flow into the cavity (and the lips are open); the cavity between B and C will have positive pressure, as in Fig. 1b, since there is continuous air flow into the cavity which has no vent.

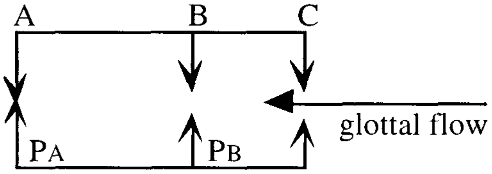


Figure 4a: labial closure, no velar closure; open glottis
$\underline{P_A} = +p$, $\underline{P_B} = +p$

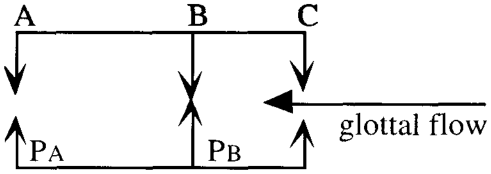


Figure 4b: no labial closure, velar closure; open glottis
$\underline{P_A} = 0$, $\underline{P_B} = +p$

Now consider simultaneous closures at points A and B in Fig. 5a. If closure at A is achieved before B, both $\underline{P_A}$ and $\underline{P_B}$ will be positive, although $\underline{P_A}$ might be lower than (maximal) $\underline{P_B}$, if closure at B is formed before glottal flow ceases. If, however, velar closure is achieved before labial closure, the pressure in the cavity between B and C will be positive while the pressure in the cavity between A and B will be at or near zero, since there is no incoming air after the formation of labial stop.

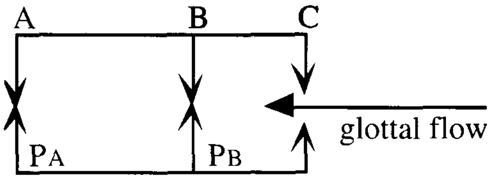


Figure 5a: labial closure, velar closure; open glottis
$\underline{P_A} = +p$ or 0, $\underline{P_B} = +p$

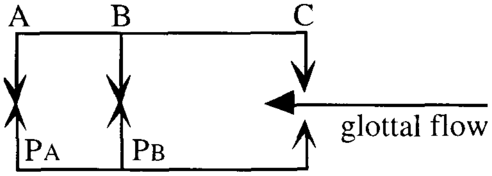


Figure 5b: labial closure, velar closure; open glottis
tongue dorsum fronting
$\underline{P_A}^{\uparrow}$, $\underline{P_B}^{\uparrow}$ (slower)

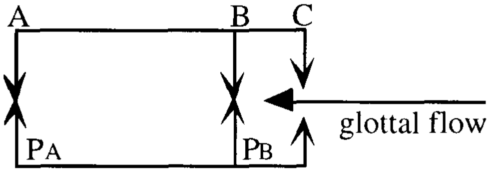


Figure 5c: labial closure, velar closure; open glottis
tongue dorsum backing
$\underline{P_A}^{\downarrow}$, $\underline{P_B}^{\uparrow}$ (faster)

Consider now the dynamic pressure changes in the two subcavities above the glottis in Figs. 5b and 5c. In Fig. 5b, the constriction at B has moved forward from the starting position in Fig. 5a, due to co-articulation of a back-front vowel sequence. This articulator movement results in a compression of the cavity sealed at points A and B; consequently, $\underline{P_A}$ increases. In the absence of glottal flow, the accompanying expansion of the cavity between B and C (due to tongue dorsum fronting) will result in pressure rarefaction, i.e., a decrease in $\underline{P_B}$ is expected. However, since Fig. 5b indicates that there *is* continuous incoming glottal flow to contribute to pressure build-up, an increase in $\underline{P_B}$ is still expected, provided the rate of pressure rarefaction due to cavity expansion is less than the rate of pressure increase due to glottal flow. Assuming this to be the case, the net effect of cavity expansion and glottal flow will result in a(n initially) slower increase in $\underline{P_B}$ than in Fig. 5a, where no cavity expansion obtains.

In Fig. 5c, the cavity sealed at constriction points A and B expands due to tongue retraction during the articulation of a flanking front-back vowel sequence. As a consequence, the enclosed air column rarifies, resulting in a decrease in $\underline{P_A}$. The reduced volume between B and C will see a faster pressure increase than would be expected from glottal air flow alone, due to the cavity compression resulting from tongue retraction.

Although I have only considered the oral pressure dynamics during labial and velar constrictions, the discussions above on multiple supralaryngeal constrictions may be extended to cover the alveolar place of articulation at constriction point B (or A).

There exist two more relevant factors to be considered: nasal opening and voicing. Although the above discussions have assumed a raised velum (i.e., no nasal leakage) and a spread glottis (i.e., considerable glottal flow) as the conditions during glottal flow, similar oral pressure consequences can be assumed during voicing (low glottal flow) and nasalization (lowered velum). However, both conditions will result in a slower supralaryngeal pressure build-up: Voicing means a dramatically reduced rate of air flowing into the oral cavity; nasalization has the same effect of reducing the net flow into the oral cavity, by the nasal vent. Less flow will result in a slower pressure build-up.

4.3.2 *Silverman & Jun (1994)*

Silverman and Jun (1994) showed that oral pressure (henceforth Po) measurements taken behind the lips during unreduced labial+velar stop sequences may serve as good diagnostics for gestural overlap when the consonant sequence is flanked by vowels of varying backness, and that the presence or absence of overlap is especially clear when the vowel environment is front-back. In ipku, an overlapped sequence will show a positive-then-negative change in Po (Fig. 4a - Fig. 5c - Fig. 4b); in ikpu, an overlapped sequence will show a negative-then-positive change in Po (Fig. 4b - Fig. 5c - Fig. 4a). In both cases, the positive change in Po is the result of pressure building up behind the labial closure, while the negative change in Po results from pressure rarefaction due to tongue retraction occurring while both the labial and velar closures are maintained (i.e., when there is articulatory overlap). Furthermore, Silverman and Jun interpreted an absence of Po change during the articulation of pk as evidence of labial reduction: In the absence of labial closure, the pressure behind the lips is expected to equal the atmospheric pressure, or zero in the metrics employed (See the discussion of Figs. 1a, 4b). Silverman and Jun (1994) were able to establish the validity of their methods because pharyngeal pressure was recorded concurrent with the Po traces; the pharyngeal pressure traces allowed them to interpret their Po data convincingly by enabling them to directly identify the duration of velar closure (as the period of positive pharyngeal pressure change).

Adopting the methods and interpretations of Silverman and Jun (1994), Jun (1995) extended the test tokens to include Korean pt clusters in addition to pk clusters; both pt and pk were preceded by the front vowel [i], and followed by a back vowel from the set [u,o,a]. In addition to being good indications of articulatory overlap during unreduced ipku, Po data is also a good diagnostic for labial reduction. In his Po trace interpretations of both pk and pt, a positive Po was taken as indication of labial closure, while a flat Po trace during the CC duration was interpreted to mean labial reduction. No conclusions were drawn from the Po data regarding the overlap of unreduced pt, based on the assumption that tongue dorsum retraction during a front-back vowel sequence will show no (significant) effect on the articulation of the coronal [t], i.e., the tongue blade will not retract along with the tongue body.

I have applied this method in assessing the presence of coda reduction in Taiwanese consonant clusters. This experiment considered non-homorganic, heterosyllabic CC sequences where either C1 or C2 was labial. In labial-initial clusters, the overall change in Po was taken to be a reliable indication of whether the coda gesture had undergone reduction. In labial-final clusters, reliable interpretation of coda reduction depended on considerable overlap of the two CC gestures, since only in this case can a meaningful interpretation of the status of the non-labial coda be inferred from the Po readings taken behind the lips; fortunately, this turned out to be the case almost all of the time.

We are now equipped with the necessary information for the understanding of the Taiwanese oral pressure experiment, to be described in the following sections.

4.4 *Method*

4.4.1 *Subjects*

Six paid native speakers (4 male, 2 female) of Taiwanese participated in this study. With the exception of one female speaker who was an undergraduate student at UCLA, all the subjects were UCLA graduate students. All of them were born and raised in Taiwan as bilingual speakers of Taiwanese and Mandarin, and left Taiwan after age 23. None of them had been abroad for more than 5 years at the time of this study, and Taiwanese is the language in the home for all of the speakers. I will refer to these speakers as "m1," "m2," "m3," "m4," "f1," and "f2," where "m1" is Male Speaker 1, and "f1" is Female Speaker 1, etc.

None of the subjects in this experiment were linguists, and all of them were naive about the purpose of the study, as well as about Taiwanese phonetics and phonology.

4.4.2 *Speech materials*

All the relevant sequences were placed in a phrase-medial position, where a fairly constant subglottal pressure (i.e., maximum potential Po) may be assumed. The phrases employed for this experiment are provided in Appendix A, with the relevant sequences underscored. There are two parts to this data set:

Part I: Word-initial versus word-final stops

It is clear from perception that word-final stops, but not word-initial stops, lenite intervocalically. Based on this assumption, Po data of intervocalic word-initial and word-final /p/ and /m/ are compared to ensure that lenition is unambiguously reflected in the Po traces. The phrases used for this part of the experiment correspond to Utterances 1~4 in Appendix A. The relevant (utterance-medial) VCV sequences are summarized in Table 1, with the word-boundary indicated by a period:

Table 1. Intervocalic Word-initial versus Word-final C

	Word-Initial	Word-Final
/p/	a.pa	ap.e
/m/	a.ma	am.e

As discussed in Section 4.3.1 in regard to Figs. 2a and 2b, glottalization will result in no (substantial) Po build-up if labial closure is achieved after glottal closure. Since no oral flow data is recorded (which would allow us to ascertain the presence of labial reduction plus an open glottis, in which case positive oral flow will be detected) when zero Po is detected, we must rely on the acoustics: Presence of voicing may be used as evidence against glottal closure whether or not the supra-laryngeal constriction is reduced.

The Po effect is expected to be more robust for /p/ than for /m/, since the positive Po build-up during /m/ will be low, if at all, due to the lower rate of glottal flow during voicing and the presence of the nasal vent which reduces the net flow into the oral cavity. It is therefore important to test what kinds of Po patterns can be used reliably as the diagnostic for the presence or absence of labial reduction.

Part II: Labial + C Clusters

Po data of the utterance-medial sequences listed in Table 2 are examined for evidence of gestural reduction of C1. Due to the limitations of the methodology, only non-homorganic stop clusters containing a labial constituent are employed. Both oral and nasal word-final stops are included in the data set. These correspond to Utterances 5 ~ 12 in Appendix A.

Table 2. Labial + Non-labial Clusters

	C1 = Labial		C2 = Labial	
	C2 = velar	C2 = coronal	C1 = velar	C1 = coronal
C1 = oral	iPku	iPtu	iKpu	iTpu
C1 = nasal	imku	imtu	iŋpu	inpu

4.4.3 *Data-collecting procedures*

Each subject was fitted with a Rothenberg mask connected to pressure/flow transducers. The pressure tube consists of two segments: The first part is 5 3/4 inches (≈ 14.5 cm) in length, and has a 5 mm inside diameter; this is connected to the pressure transducer at one end, and to the second tube at the other. The second part is 3/4 inches (≈ 2 cm) in length, and has a 3 mm inside diameter. The smaller-diametered tube is used to minimize the disruptiveness of the experimental apparatus for the subjects. The longer, larger-diametered tube is required by the design of the mask and the transducer; it also gives a better frequency response than a narrower tube would. The tubes, after being connected together, measure about 6 inches in length.

The subjects were asked to hold the mask, and insert the narrower end of the tube in the corner of the mouth between the lips[19]. They were instructed to not allow the end of the tube to come in contact with their teeth. When the smaller tube is held between the lips, its (opening) end is probably perpendicular to the major direction of flow; oral pressure measured in this way reduces the potential for artifacts in pressure record.

The subjects were provided with twelve Taiwanese phrases, read to them one at a time by the experimenter. After the experimenter read each phrase, the subject was instructed to repeat the phrase eight times in their normal, conversational speech style.

Two signals were recorded on the Kay Elemetrics Computerized Speech Lab (CSL) in the UCLA Phonetics Laboratory: audio and oral pressure (behind the lips). The first and eighth repetitions of each test phrase were discarded. Flow[20] was not captured because a tight fit of the mask could not be achieved for any of the speakers.

The oral pressure trace was not calibrated because we are interested here only in the qualitative pattern of Po change. Also, we present below the unfiltered Po trace; this allows us to identify the periods of voicing as the regular oscillations in the Po waveforms.

4.5 *Results and analysis*

4.5.1 *Word-final oral stops*

Part I: Intervocalic consonants

The Data

Figs. 6a and 6b provide representative Po traces of intervocalic word-initial /p/ and word-final /P/. All the tokens (6 repetitions per token x 6 speakers) examined displayed the same pattern.

In each figure, the top waveform corresponds to the audio signal, and the bottom is the Po trace. The vertical axis represents pressure, while the horizontal axis represents time in seconds. A phonemic transcription for the utterance is provided below the audio waveform; morpheme boundaries are indicated by a period ".". I have marked off the relevant segment (underlined in the transcription) in each waveform by the vertical lines labeled A and B. Point A corresponds to the acoustic onset of the consonant, and was located from an expanded waveform by marking the end of the vowel (complex wave). Point B corresponds to the release (offset) of the consonant, and was located from the waveform by marking the stop burst in the case of unreduced consonants, or by marking the beginning of the following vowel (complex wave in the case of a reduced labial).

---

[19] See Jun 1995: 99, Fig. 2 for a diagramatic representation of the method of measuring oral pressure.

[20] As mentioned earlier, oral flow data is helpful in the word-final, intervocalic position in disambiguating between glottalization and labial reduction with no glottalization, both of which will result in no pressure build-up behind the lips, but only the latter will yield positive oral flow. In addition, flow data is helpful in identifying the interval of oral closure in consonant clusters (= the interval of no oral flow).

Fig. 6a contains two instances of Po rise during an intervocalic word-initial /p/[21]; I have only marked off one instantiation below:

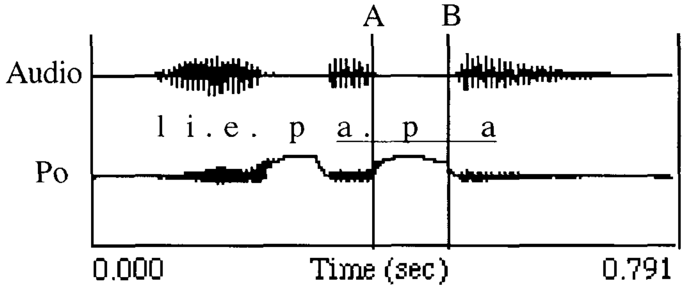


Figure 6a: Word-initial /p/ in a.pa

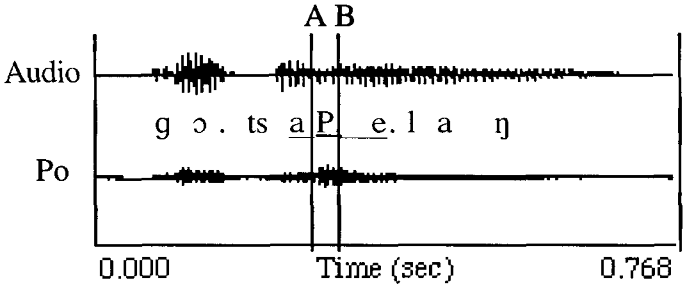


Figure 6b: Word-final /P/ in aP.e

Figs. 6a and 6b are representative of the intervocalic /p/ and /P/ examined. We observe from a comparison of Figs. 6a and 6b that positive Po is detected during intervocalic word-initial /p/, where no reduction is perceived, while no overall Po rise is detected during intervocalic word-final /P/, where labial reduction is expected. An additional observation is that word-initial /p/ is significantly longer than word-final /P/.

Notice that in Fig. 6b, although a voiceless closure presumably obtained during the affricate /ts/, no positive Po rise is detected. This is because the current method will only detect Po (immediately) behind the lips; the Po accruing behind the coronal constriction is not detectable by the present methodology.

Discussion

Po data unambiguously showed that consistently across 6 speakers, word-initial /p/ was unreduced while word-final /P/ was reduced: A significant pressure rise was detected during the production of word-initial /p/ (Fig. 6a), as predicted in the discussion of Fig. 1b, while no (significant) overall Po change was detected during word-final /P/ (Fig. 6b), as predicted in the discussion of Fig. 1a. Furthermore, the intervocalic Final /P/ duration showed sinusoidal oscillation on the audio waveform indicative of voicing; the absence of Po rise is therefore not the consequence of glottalization (i.e., *not* the scenario diagrammed in Figs. 2a or 2b). Finally, although it might be argued that the lack of positive Po in /P/ is *not* due to lenition, but is due to a combination of reduced glottal flow (voicing) plus short duration, it is clear from the percept (roughly, IPA [β]) that intervocalic /P/ was always reduced. Thus, in interpreting Po of CC clusters, *I will consider a positive Po similar to that in a.pa as indication of unreduced labial closure, while an absence of (significant) overall Po change during /P/ or /p/ (accompanied by voicing) such as in ap.e. as indicating labial reduction.*

---

[21] I consider the string /pa pa/ 'Dad' to be a reduplicated form of the vocative /pa/ 'Dad,' which is also well-formed. Both /p/s in /pa pa/ are therefore word-initial.

Part II: Consonant clusters

Summary of attested Po patterns during CC

I use the following notations to qualitatively categorize the Po traces:

AP: All-Positive; positive Po was detected throughout the CC duration.

PN: Positive-then-Negative; positive Po was detected, followed by negative Po.

PZ: Positive-then-Zero; positive Po was detected during (roughly) the first half of the CC duration, followed by zero Po during the remaining CC duration.

AZ: All-Zero; no change in Po (zero) was detected throughout the CC duration.

Figs 7 to 10 exemplify these Po patterns. As in the previous figures, the top waveform in each figure below is the audio, and the bottom is the oral pressure measured behind the lips. The vertical axis measures pressure, while the horizontal axis measures time. The vertical lines labeled A and B demarcate the duration of the consonant cluster CC, where either C1 or C2 is a labial consonant. Point A is determined from the expanded waveform as the point where the vowel [i] (complex waveform) ends. Point B corresponds to the burst of C2, or the release of the stop closure. A phonemic transcription is provided below the audio waveform; the relevant consonant cluster is underlined, and the morpheme boundaries are indicated by a period ".".

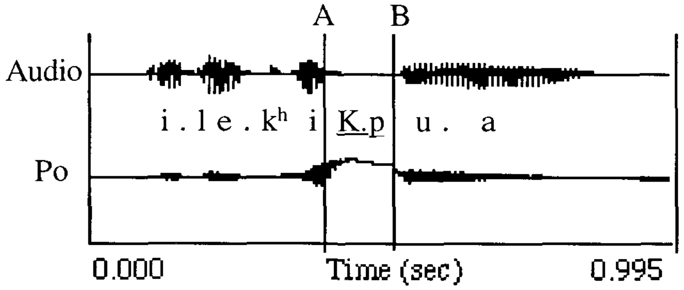


Figure 7: AP (All-Positive)
/iK.pu/. The example is taken from Speaker m1.

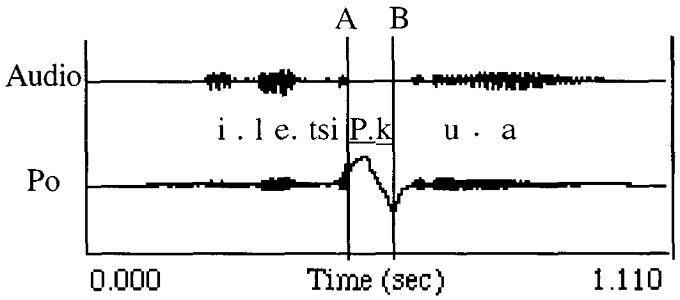


Figure 8a: PN (Positive-then-Negative)
/iP.ku/. The example is taken from Speaker m1.

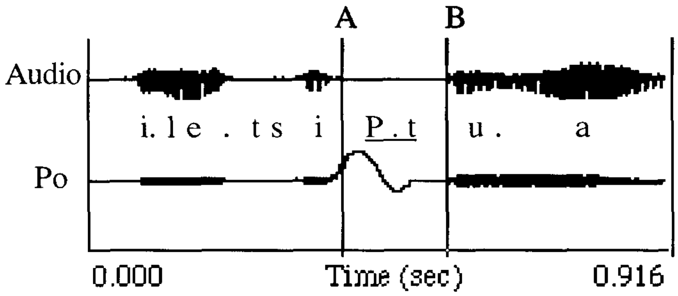


Figure 8b: PN (Positive-then-Negative)
/iP.tu/. The example is taken from Speaker f1.

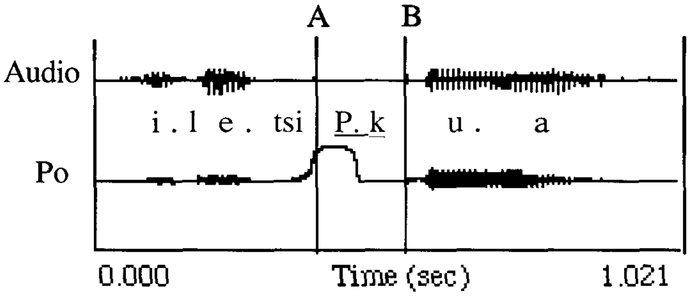


Figure 9: PZ (Positive-then-Zero)
/iP.ku/. The example is taken from Speaker m4.

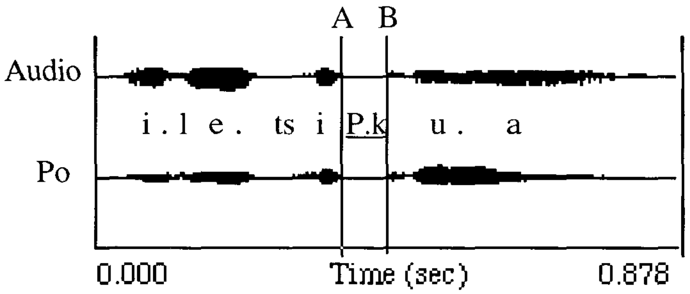


Figure 10: AZ (All-Zero).
/iP.ku/. The example is taken from Speaker f2.

Table 3 summarizes the number of times (out of 36 repetitions) each pattern was attested during the consonant sequences (flanked by the vowels i-u) listed in the far left column. The patterns are indicated on the first row, not in any significant order. To facilitate the reading of this table, I have shaded the boxes for which zero results were obtained. Thus, the remaining unshaded boxes report only positive (non-zero) results.

Table 3. Summary of Results by Po Patterns

	AP	PN	PZ	AZ
iP.ku		15	12	9
iP.tu		5	23	8
iK.pu	36			
iT.pu	36			

Appendix B presents the (same) results by speaker.

Interpretation of the Po data

The results in Table 3 are interpreted based on the following assumptions:

i. An AP pattern indicates complete labial closure throughout the entire duration of CC. Moreover, no other (significant) consonantal constriction was formed during the labial closure (See discussion of Fig. 4a in Section 4.3.1). In the case of a labial C1, this would mean progressive place assimilation by the labial C1 plus gestural reduction of the non-labial C2. In the case of a labial C2, this would mean regressive place assimilation by the labial C2 plus gestural reduction of the non-labial C1. An alternative interpretation is gestural overlap of CC plus lenition of the non-labial consonant. As we might expect, this pattern never arises in the labial-C1 clusters, i.e., word-initial lenition plus compensatory lengthening by the preceding coda is unattested in Taiwanese.

ii. As discussed in Silverman and Jun (1994), a PN pattern indicates simultaneous closure by a labial C1 and a non-labial C2, or, gestural overlap plus the absence of lenition at both places of articulation; the rarefaction is due to tongue body retraction which expands the sealed cavity (See discussion of Fig. 5c in Section 4.3.1). Thus, the pattern is attested only in labial-C1 clusters. In addition, as exemplified in Figs. 8a and 8b, both iP.ku and iP.tu may yield this pattern, contrary to the assumption in Jun (1995) that front-back vowel co-articulation on an overlapped labial-coronal cluster will have no significant effect (of pressure rarefaction) on the Po measured behind the lips.

iii. A PZ pattern indicates labial closure during the first part of the CC duration only. Furthermore, no significant gestural overlap by the non-labial occurred during labial closure. Not surprisingly, this pattern is only attested in the labial-C1 clusters.

iv. An AZ pattern indicates that no labial closure was formed throughout the CC closure. In the case of no gestural overlap, this would either mean a lenited labial C1 or a lenited labial C2 (Fig. 1a). Although no Po will be detected behind the lips if the labial closure is formed after the non-labial closure (i.e., [kp] or [tp]), all else being equal, the design of the present experiment will yield a NP Po pattern due to co-articulation of the front-back flanking vowels. It is thus not surprising that this pattern is attested only for the labial-C1 clusters, since, based on perception, we do not expect to find evidence of word-initial reduction.

To summarize, the results in Table 3 are interpreted based on the following criteria:

i. C1 is interpreted as unreduced in P.C2 if either positive-then-zero (PZ: no overlap of Pk, Fig.9) or positive-then-negative (PN: overlap of P.C2, Figs. 8a,b) Po patterns were detected during the CC duration. C1 is interpreted as reduced in P.C2 (Fig. 10) if no change in Po (AZ) was detected.

ii. For Pt sequences, no conclusion is drawn regarding the degree of gestural overlap from an absence of negative pressure, assuming, according to Jun (1995), that rarefaction will not necessarily be detectable in overlapped, unreduced iP.tu sequences.

iii. C1 reduction is interpretable in C1.p only when C1.p was highly overlapped, and a non-zero Po was detected throughout the duration of C1.p. In such case, an all-positive (AP) Po pattern is interpreted as indicating gestural reduction in C1 (Fig. 7), while a negative-then-positive (NP) Po is interpreted as no reduction in C1 (Fig. 12); the latter pattern never arose in the oral C1.p clusters.

Summary of reduction in C1 and gestural overlap in CC

The numbers indicated in Tables 4a and 4b are derived from the results summarized in Table 3. The denominators in each cell indicate the total number of unambiguous repetitions from the experiment; the maximum number is 36 in each case (6 speakers x 6 repetitions for each token). The sum of the denominator (counted once for each token) and the number of inconclusive tokens should equal 36. In the tables below, I have shaded the boxes for which no unambiguous data may be obtained. The reasons for the choice of the shaded box(es) are as follows:

i. iP.tu, no overlap: Reliable evidence for the non-overlap of Pt is obtainable only when the labial is unreduced (since both an overlapped and unoverlapped [βt] sequence are expected to register zero pressure behind the lips). Assuming this to be the case, non-overlapped Pt should exhibit a positive-then-zero (PZ) Po trace. However, Jun (1995) suggested that the same Po trace may result from an overlapped Pt in the i-u environment, since it is not clear that the tongue-body retraction need necessarily result in the rarefaction of the air in the cavity sealed by the lips and the tongue blade[22], assuming that the tongue blade will not always retract in a front-back vowel environment. Although some instances of PN patterns have been obtained during iP.tu sequences, thereby providing unambiguous evidence of gestural overlap, in the absence of pharyngeal pressure data, I will assume that the PZ Po pattern during iP.tu has an ambiguous interpretation between non-overlap and overlap.

Table 4a: Word-final (C1) Reduction in CC

	iP.ku	iP.tu	iK.pu	iT.pu
reduction	9/36	8/36	36/36	36/36
no reduction	27/36	28/36	0/36	0/36
inconclusive	0	0	0	0

---

[22] That negative pressure was obtained at all for labial-t sequences (5 for iptu, 16 for imtu) suggests one of three possible mechanisms of cavity expansion (pressure rarefaction):

i. The tongue blade retracts along with the tongue body due to vowel co-articulation.

ii. The tongue blade does not retract, but the retraction of the tongue body "stretches" out the anterior part of the tongue, causes the curvature of the blade to go from a concave down configuration to a concave up configuration, thus expanding the cavity sealed by the tongue tip and the lips.

iii. The lips protrude in anticipation of the rounded vowel to follow (suggested by P. Keating). The pressure rarefaction may reflect one of these articulatory possibilities, or any combination thereof.

I infer from Table 4a that word-final reduction may occur for all places of articulation. Moreover, non-labials are more likely to reduce than labials. More specifically, in the present data corpus, labials reduce before [k, t] about 25% of the time, and non-labials reduce before [p] 100% of the time.

Table 4b: Gestural Overlap in CC

	iP.ku	iP.tu	iK.pu	iT.pu
overlap	15/27	5/5	36/36	36/36
no overlap	12/27		0/36	0/36
inconclusive	9	31	0	0

I infer from Table 4b that both labial-C1 and labial-C2 clusters may overlap. Moreover, I generalize that labial-C2 clusters are more likely to overlap than labial-C1 clusters. More precisely, the present corpus offers unambiguous evidence of no overlap in some tokens of iP.ku and unambiguous evidence of overlap in some tokens of iP.ku and iP.tu; all instances of iK.pu and iT.pu are interpreted to be (unambiguously) overlapping.

Conclusion

This portion of the experiment supports the hypothesis that C1 may reduce in CC. Keeping in mind that C2 is not constant in these test tokens, we can also generalize that either labials reduce less readily than velars and coronals, or that labials are more likely than velars and coronals to trigger reduction of the preceding C, or both. Put another way, this experiment has established that it is more likely for non-labials to reduce before labials than for labials to reduce before non-labials in Taiwanese (stop) consonant clusters. Note that the present oral pressure experiment is unable to show directly that Labials are less likely to reduce than Velars, since the design of the experiment cannot provide evidence of the tongue dorsum activity in a CC sequence where C1 is /k/ and C2 is /t/[23]. Also, the oral pressure experiment will only detect consonantal weakening if the extent of the weakening results in the absence of stop closure. In this view, it is possible that the degree of lenition is the same at all places of articulation, but "more" lenition needs to obtain in labials than in non-labials for reduction to be reflected in Po data, since the Po data can only distinguish between presence and absence of closure.

In addition, at least with respect to word-final labials, the results of this experiment showed that while labial reduction *sometimes* occurs before a following (non-homorganic) stop, lenition *always* obtains before a following vowel.

This experiment also showed that in Taiwanese, gestural overlap in non-homorganic consonant sequences is more common in labial-final clusters than in labial-initial clusters.

To conclude, the results of this experiment support the prediction based on the attested place assimilation and Jun's (1995) conclusion that gestural reduction is responsible for the perceptual loss of the target place in casual speech place assimilation: Word-final stops may reduce before a following consonant. The hypothesis that non-word-initial lenition is a more general process, occurring before vowels and consonants alike, is supported.

Section 4.4.2 is a similar experiment which assesses word-final reduction in nasals. Similar conclusions regarding reduction and overlap are reached.

---

[23] Similarly, while electropalatography (EPG) will be able to shed light on the tongue dorsum activity in the sequence K.t, it cannot also provide comparable information on the labial activity in the sequence P.t. An EPG study, or a combination of EPG and Po study will thus still be unable to establish the relative likelihood of word-final /P/ reduction and /K/ reduction, since while EPG data may provide gradient evidence of articulatory reduction, Po data will only provide categorical evidence of (continuant-yielding) articulatory reduction.

4.5.2 *Word-final nasal stops*

Part I: Intervocalic consonants

The Data

Figs. 11a and 11b provide representative Po traces of intervocalic word-initial and word-final /m/. All the tokens (6 repetitions per token x 6 speakers) examined displayed the same pattern.

In each figure, the top waveform corresponds to the audio signal, and the bottom is the Po trace. The vertical axis represents pressure, while the horizontal axis represents time. A phonemic transcription for the utterance is provided below the audio waveform; morpheme boundaries are indicated by a period ".". I have marked off the relevant segment (underlined in the transcription) in each waveform by the vertical lines labeled A and B. Point A corresponds to the acoustic onset of the consonant, and was located from an expanded waveform by marking the end of the vowel (complex waveform) duration. Point B corresponds to the release of the consonant, and was located from the waveform by marking the beginning of the following vowel (complex waveform).

Fig. 11a contains two instances of Po rise during an intervocalic word-initial /m/; I have only marked off one instantiation below:

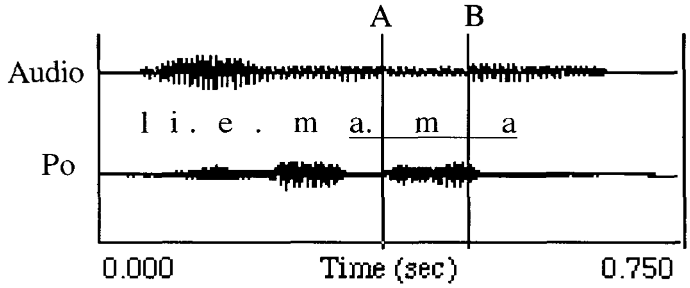


Figure 11a: Word-initial /m/ in a.ma

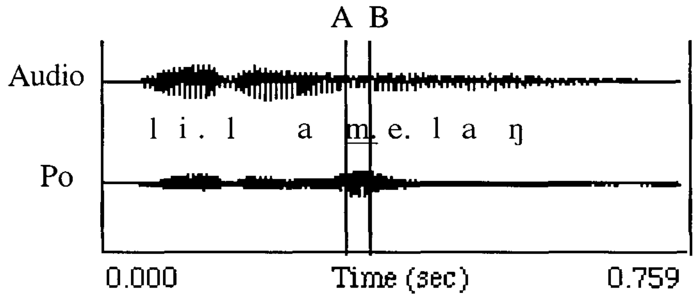


Figure 11b: Word-final /m/ in am.e

Figs. 11a and 11b are representative of the intervocalic word-initial and word-final /m/ examined. We observe from a comparison of Figs. 11a and 11b that no significant Po rise is observed in either word-initial or word-final intervocalic /m/, even though labial reduction is expected only in the latter. An additional observation is that word-initial /m/ is significantly longer than word-final /m/; the same contrast in segment duration was found in the oral consonants.

Discussion

In contrast to the oral consonants, Po data is unable to distinguish the degree of oral constriction/reduction between intervocalic word-initial /m/ (Fig. 11a) and word-final /m/ (Fig. 11b); both showed no (or very slight) overall Po change, consistent across 6 speakers. Assuming the labial closure was complete in word-initial /m/ (as in word-initial /p/), the absence of positive Po during /m/ is likely due to a combination of the voicing configuration of the glottis (i.e., less air flowing into the vocal tract) and nasal leakage. I conclude that this lack of asymmetry between the Po data of word-initial versus word-final /m/ makes m.C2 clusters (or more generally, clusters containing a nasal coda) less appropriate for the current investigation. I have nonetheless included NC clusters in the data set since any instance of positive (or negative) Po during the nasal duration must necessarily be interpreted as indicating an unreduced labial closure, and, in the case of labial C2 clusters, temporal overlap of NC constrictions plus a reduction of N. In interpreting Po of CC clusters, *I will consider a non-zero Po as indication of unreduced labial closure, while an absence of Po change during the duration of /m/ is inconclusive evidence for labial reduction.*

Part II: Consonant clusters

Summary of attested Po patterns during CC

In addition to the notations AP, PN, PZ, and AZ, introduced in Section 4.5.1, the following notations are used in the qualitative categorization of the Po traces in this section:

NP: Negative-then-Positive; negative Po was detected, followed by positive Po.

ZP: Zero-then-Positive; zero Po was detected during the first half of the CC duration, followed by positive Po during the remaining CC duration.

ZN: Zero-then-Negative; zero Po was detected during the first half of the CC duration, followed by negative Po during the remaining CC duration.

Figs. 12, 13, and 14 instantiate these Po patterns. In each figure below, the top waveform is the audio, and the bottom is the oral pressure measured behind the lips. The vertical lines labeled A and B demarcate the duration of the consonant cluster CC, where either C1 or C2 is a labial consonant. Point A is determined from the expanded waveform as the point where the vowel [i] (complex waveform) ends. Point B corresponds to the onset of the vowel [u] (complex waveform). A phonemic transcription is provided below the audio waveform; the relevant consonant cluster is underlined, and the morpheme boundaries are indicated by a period ".".

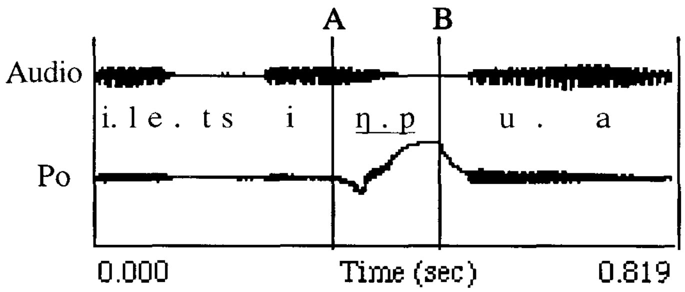


Figure 12: NP (Negative-then-Positive)
/iŋ.pu/. The example is taken from Speaker f1.

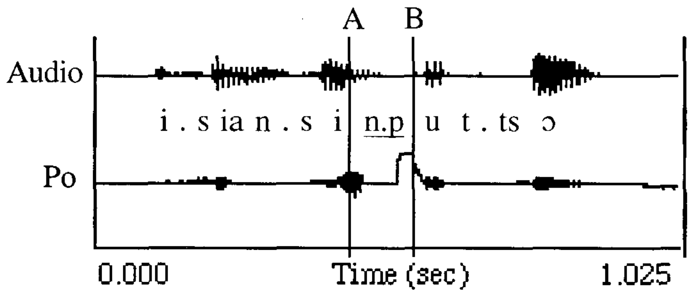


Figure 13: ZP (Zero-then-Positive)
/in.pu/. The example is taken from Speaker m2.

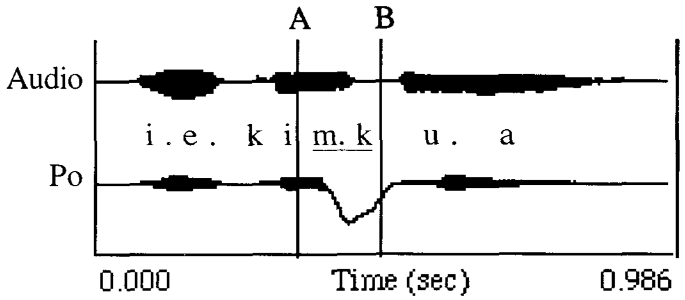


Figure 14: ZN (Zero-then-Negative)
/im.ku/. The example is taken from Speaker f2.

There are also instances of Po rise (AP, PN, PZ) during the word-final nasal. However, unlike the Po rise during oral consonants, the rise is more gradual during nasal consonants, a likely consequence of the reduced glottal flow and nasal leakage. An example of AP is provided in Fig. 15. Notice that the peak in the Po trace is only reached in the latter part of the cluster duration, upon the offset of (nasal) voicing and the onset of the voiceless stop (cf., Fig. 7: peak Po reached during the first half of K.p).

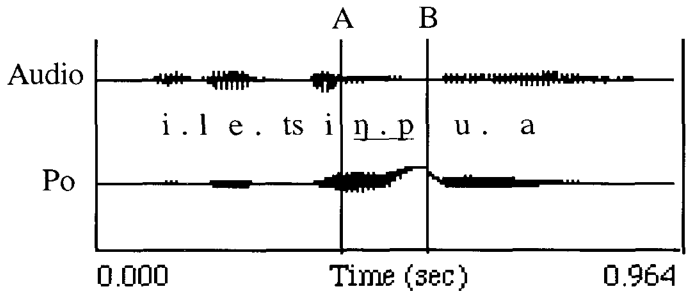


Figure 15: AP (All-Positive)
/iŋ.pu/. The example is taken from Speaker m1.

Table 5 summarizes the number of times (out of 36 repetitions) each pattern was attested during the consonant sequences (flanked by the vowels i-u) listed in the far left column. The patterns are indicated on the first row, not in any significant order. To facilitate the reading of this table, I have shaded the boxes for which zero results were obtained. Thus, the remaining unshaded boxes report only positive (non-zero) results.

Table 5. Summary of Results by Po Patterns

	AP	PN	NP	PZ	ZP	ZN	NC
im.ku		26		2		4	4
im.tu		16		10			10
iŋ.pu	30		5		1		
in.pu	31				5		

Appendix B presents the (same) results by speaker.

Interpretation of the Po data

The results in Table 5 were interpreted based on the following assumptions:

i. As discussed in Silverman and Jun (1994), a NP pattern indicates simultaneous closure by a non-labial C1 and a labial C2, i.e., there is gestural overlap and no reduction.

ii. A ZP pattern may indicate either labial closure during the second part of CC only, plus no significant gestural overlap, or labial closure during the entire duration of CC plus reduction of the non-labial (C1). Recall that we have concluded in Part I that a labial closure during nasal voicing will not necessarily result in a Po rise.

iii. A ZN pattern indicates simultaneous closure by a labial C1 and a non-labial C2, i.e., there is articulatory overlap plus no reduction; I consider this to be a subcase of PN. The ZN pattern further validates my conclusion from Part I that a zero change in Po during a nasal segment is *not* a reliable indication that the labial gesture is reduced, since the negative Po indicative of pressure rarefaction can only obtain in a sealed cavity.

iv. The patterns AP, PN, PZ, and AZ are interpreted in the same way as in Section 4.5.1.

To summarize, the results in Table 5 are interpreted based on the following criteria:

i. C1 is interpreted as unreduced in m.C2 if either positive-only (PZ: no overlap of mk) or positive-then-negative (PN: overlap of mC) Po were detected during the CC duration; the ZN cases are treated as an extreme subcase of PN. No conclusion is drawn regarding the status of labial reduction or gestural overlap from a zero change (AZ) in Po.

ii. For mt sequences, no conclusion is drawn regarding the degree of gestural overlap from an absence of negative pressure.

iv. C1 reduction is interpretable in C1.p only when C1.p was highly overlapped, and a non-zero Po was detected throughout the duration of C1.p. In such case, a positive-only (AP) Po pattern is interpreted as indicating gestural reduction in C1, while a negative-then-positive (NP) Po is interpreted as no reduction in C1.

v. C1 reduction is interpretable in Np only when Np was highly overlapped. In such case, a positive-only (AP) Po during CC is interpreted to mean gestural reduction in C1, while a negative-then-positive Po (NP) is interpreted to mean no reduction in C1. For Np sequences, initial zero change in Po (e.g., ZP) is considered an inconclusive indication of reduced C1 coupled by overlapping /p/, since this scenario cannot be distinguished from the non-overlapping CC scenario.

Summary of reduction in C1 and gestural overlap in CC

The numbers indicated in Tables 6a and 6b are derived from the results summarized in Table 5. Again, the total number of repetitions of each token is 36 (6 speakers x 6 repetitions per token); the denominators indicate the number of repetitions showing conclusive evidence of (non-)reduction/(non-)overlap. I have shaded the boxes for which no unambiguous data may be obtained. The reasons for the choice of the shaded boxes are as follows:

i. im.ku, im.tu, no reduction: The expected Po for reduced /m/ would be one in which there is no overall change in Po (AZ). However, we have concluded from Part I that such a Po trace is also obtainable from an unreduced [m].

ii. im.tu, no overlap: Reliable evidence for the non-overlap of labial-C is obtainable only when the labial is unreduced. Assuming this to be the case, non-overlapped labial-t should exhibit a positive-then-zero (PZ) Po trace. However, I have concluded in

Section 4.4.1 that the same Po trace may result from an overlapped labial-t in the i-u environment.

iii. iŋ.pu, in.pu, no overlap: The expected Po trace for a non-overlapped nasal-p is zero-then-positive (ZP). However, the same trace may result from a highly overlapped Np with a reduced C1. Since the first portion of this sequence is produced with velic opening, and probably also voicing, a positive Po will not necessarily be detected.

Table 6a: Word-final Reduction in CC

	im.ku	im.tu	iŋ.pu	in.pu
reduction			30/35	31/31
no reduction	32/32	26/26	5/35	0/31
inconclusive	4	10	1	5

I infer from Table 6a that word-final labial nasals are less likely to reduce than non-labial nasals. Furthermore, non-labial nasals frequently reduce before [p].

Table 6b: Gestural Overlap in CC

	imku	imtu	iŋpu	inpu
overlap	30/32	16/16	35/35	31/31
no overlap	2/32			
inconclusive	4	20	1	5

I infer from Table 6b that word-final nasals frequently overlap with a following consonant. Furthermore, it is more likely for a non-labial C1 and a labial C2 to display gestural overlap than for a labial C1 and a non-labial C2 to show overlap.

Conclusion

This portion of the experiment supports the hypothesis that word-final nasals as well as oral stops may undergo reduction in the pre-consonantal position[24]. Furthermore, the observation from Section 4.5.1 that labial reduction is less frequent than non-labial reduction is corroborated by the NC sequences.

In addition, we have shown that the generalization that gestural overlap obtains in non-homorganic clusters, and that the overlap is more common in labial-C2 clusters than in labial-C1 clusters, also applies to NC sequences.

4.6 *Conclusion*

This experiment supports the hypothesis that Taiwanese pre-consonantal codas may be lenited in casual speech. Moreover, we have found that word-final reduction is optional before a consonant, and that pre-consonantal reduction is less frequent than pre-vocalic reduction.

Furthermore, results of this experiment suggest that (continuant-yielding) non-labial reduction is more common than labial reduction, and that overlap by a labial C2 is more common than overlap by a non-labial C2[25]. These generalizations have been shown to apply to both oral and nasal consonants occurring in the word-final position[26].

---

[24] We might suspect that nasals lenite intervocalically as well, but that has no (significant) auditory or Po effect. One possible source of evidence may come from oral flow data, which this experiment did not collect.

[25] Although gestural overlap is not the focus of this investigation, we suspect that overlap by a labial C2 is more common than overlap by a non-labial C2 because labials are intrinsically longer than non-labials. Thus, the longer

Although the present experiment only provided data for labial + non-labial clusters, Electropalatography (EPG) studies on Taiwanese consonant clusters currently in progress by Shu-hui Peng indicate that word-final reduction also occurs in T.k and K.t clusters as well as in intervocalic T and K (Shu-hui Peng, p.c.).

Having established the occurrence of pre-consonantal reduction, we now return to the analysis of word-final lenition in Taiwanese.

5. Analysis of Consonant Reduction in Taiwanese

5.1 *Observations*

To recapitulate, this paper proposes to provide a unified account of the following observations about Taiwanese, within the framework of Optimality Theory (OT):

18 a. Word-final stops never have audible bursts, a fact which can be attributed to either overlap by a following consonant or to glottalization in the pre-pausal position.
b. Word-final stops at all places of articulation reduce intervocalically.
c. Word-final stops at all places of articulation may reduce pre-consonantally.
d. Word-initial stops never undergo lenition.
e. Lenition is detected more consistently in the intervocalic position than in the pre-consonantal position.
f. Lenition is not perceived pre-pausally.
g. Labial stops are less likely to reduce than non-labials.

In addition, there are two crucial differences between Korean (as discussed by Jun 1995) and Taiwanese we must explain in a unified account of reduction:

19 a. Taiwanese lenition occurs both intervocalically and pre-consonantally; Korean coda reduction is limited to the pre-consonantal position.
b. In Korean, labials and coronals reduce to the exclusion of velars. In Taiwanese, velars and coronals are more likely to reduce than labials.

5.2 *Analysis*

First, adopting Jun's (OT) reduction and preservation constraints introduced in Section 3.3, the Taiwanese facts in (18a-c) can be accounted for straightforwardly by the constraint ranking in (20):

20. **RED** >> Pres(pl(dor˺)) >> Pres(pl(lab˺)) >> Pres(pl(cor˺))

The constraint ranking in (20) indicates that unreleased stops (18a) at the velar, labial, and coronal places of articulation may undergo reduction (18b,c) since the preservation constraints for all three places are outranked by the reduction constraint RED. Moreover, since the constraint hierarchy does not restrict the environments where reduction may occur, we expect to find across-the-board reduction. This is indeed attested for word-final consonants in all phrase-/utterance-medial

---

labial C2 is more likely to significantly overlap with C1 than a non-labial C2 would; similarly, the longer labial C1 is less likely to be completely overlapped by C2 than a non-labial C1 would.

[26] We suggest that the results of this experiment (that K, T reduction is more common than P reduction) are unlikely to be a word-frequency effect, i.e., the consonants found to reduce more readily are not necessarily the most frequently-used lexical items. Specifically, it is my impression that kʰiK 'carve' is *not* a frequent lexical item; nevertheless, reduction of its final K was always detected during the experiment.

positions, both intervocalically and pre-consonantally. See the discussion of language L1 in Section 3.3.1.

Next, assuming a least-effort principle of speech production, languages will try to satisfy the reduction constraint RED whenever possible, that is, when there is no higher-ranked constraints which conflict with the output of RED. In Taiwanese, the presence of one such constraint is suggested by the failure of word-initial stops to undergo spirantization (18d):

21. Pres(F(w.i.)):
All features associated to the word-initial position must be preserved, where word refers to monosyllabic free morphemes[27].

In addition to ensuring a faithful realization of the stricture features of word-initial consonants (stops and continuants alike), Pres(F(w.i.)) requires that place features and other manner features (e.g., [voice], [aspiration], [sibilant]) be preserved as well. Ranking Pres(F(w.i.)) above RED for Taiwanese will capture the observation that word-initial stops do not lenite:

22. Pres(F(w.i.)) >> **RED** >> Pres(pl(dor˺)) >> Pres(pl(lab˺)) >> Pres(pl(cor˺))

Focusing on the partial constraint hierarchy underscored in (22), the tableaux in (23a) and (23b) evaluate the relevant candidates illustrating the eventual outcome of word-initial /p/ versus non-word-initial /P/ in the intervocalic position; similar outcomes (word-final lenition, word-initial preservation) are predicted for consonant clusters. In (23a,b), reduction is indicated by the lowering diacritic [˕]. We will assume that other constraints which prevent further reduction of the word-final consonant (or segmental deletion in general) are also active. For example, a general preservation constraint will ensure that some degree of labial constriction will be realized, thereby preventing deletion of the labial gesture[28] (i.e., VC.V -> *VØV).

23 a. Word-Initial /p/: No lenition.

/a.paʔ/ 'Dad'	Pres(F(w.i.))	RED
→ apaʔ		*
ap̞aʔ	!*	

b. Word-Final /P/: Lenition.

/aP.a/ 'box'	Pres(F(w.i.))	RED
aPa		!*
→ aP̞a		

In the tableaux in (23), the input is indicated in the top left cell, and the word-boundary is marked by a period. The relevant candidates are listed in the leftmost column, while the constraints are listed on the top row according to the hierarchical order in (22). Each asterisk "*" records one violation of the given constraint, and an exclamation point "!" precedes the asterisk which encodes the fatal violation. The arrow "→" has been placed next to the winning candidate form.

Returning our attention to the intervocalic labials in (23), consider first the word-initial /p/ in (23a): The second candidate satisfies RED, but commits one violation of Pres(F(w.i.)) in doing so, by failing to preserve the closure feature. In contrast, the winning candidate in (23a) with

---

[27] Recall our assumption in Section 2.1 that all morphemes are monosyllabic in Taiwanese. Furthermore, only suffixes are bound morphemes. Thus, the current definition of word includes all monosyllabic morphemes except for affixes.

[28] In Jun (1995), gestural deletion results from a different reduction constraint DELete.

respect to the medial /p/ is the unreduced [p] in [a paʔ]. Although it commits one violation of RED by failing to spirantize, it does not commit the higher-ranked Pres(F(w.i.)), which its competing candidate violates.

(23b) evaluates the relevant candidates for word-final /P/. The first candidate in (23b) retains full closure, and violates the reduction constraint. The second candidate satisfies the reduction constraint by spirantization, and emerges as the winning candidate. The constraint Pres(F(w.i.)) is irrelevant here since /P/ is not word-initial.

Before continuing our discussion of the remaining observations in (18) and (19), we must address the issue of whether the descriptive statements about Taiwanese in (18b,c,d) are best analyzed as initial preservation plus final lenition (as suggested by (22)), or as initial fortition plus final non-fortition. The simplest kind of initial fortition analysis can be ruled out immediately: Fortition, if it applies generally to the word-initial position, would generate a word-initial consonant inventory consisting entirely of stops. However, as the Taiwanese consonant inventory in (2a) indicates, both stops and continuants (s, h) occur contrastively in the word-initial position. The variable contrastive continuancy in the word-initial position thus points against an analysis of initial fortition.

A more subtle argument in favor of word-initial preservation and final lenition is that such an analysis would predict more invariant articulations in the word-initial position, where an invariant output must be preserved, but more variable[29] articulation in the non-word-initial position, where the output is dependent on the extent of lenition. This is indeed attested: Word-initial stops (and s, h) are realized as such, while non-initial consonants have variable, non-contrastive realizations in terms of the stricture degree, ranging from non-continuants to continuants. The alternative analysis of word-initial fortition would predict more variable stricture realizations initially, depending on the extent of fortition; this is not attested. For example, word-initial /p/ is always realized as the stop [p], and never as the fricative [ɸ]. Thus, we may conclude that the preservation hypothesis encoded in the constraint Pres(F(w.i.)) correctly predicts the non-spirantization of word-initial stops (as well as the continuancy preservation in /s/ and /h/). However, this is not to say that word-initial fortition may never occur. For example, it is plausible that word-initial stops are "strengthened" utterance-initially either by increasing the contact (closure) area between the articulators or by prolonging the closure duration, or both[30].

Having settled the appropriateness of the constraint Pres(F(w.i.)), we now return to the observations in (18) and (19). First, consider the Taiwanese-Korean asymmetry stated in (19a): While Taiwanese lenition occurs both intervocalically and pre-consonantally, Korean reduction is limited to pre-consonantal codas. (24) provides examples of Korean reduction (transcribed as place assimilation pre-consonantally) and non-reduction for word-final /p^h/ and /p/[31] (Sun-Ah Jun, p.c.):

---

[29] By "variable articulation," I mean variations in the articulation which may lead to *potentially contrastive* percept. For example, the velar stop [k] and the velar fricative [x] are "variable articulations" in the present sense, since the difference in their stricture degree are sufficiently different in their acoustic consequences to be contrastive in some languages, e.g., Mandarin Chinese.

[30] This has been shown for English by Fougeron and Keating (1995).

[31] In (24a), the aspirated /p^h/ is realized as an unreleased lenis stop due to a regular process of coda neutralization in Korean. In addition, the superscript ['] marks the preceding stop as fortis; this is the output of a regular process of post-obstruent fortition in Korean. In (24b), the gloss *NOM* stands for the Nominative Case marker.

24 a. Pre-consonantal stop: optional lenition
/ip^h + kwa/ -> [ipk'wa] ~ [ikk'wa]
'leaf' and'

/ip + kwa/ -> [ipk'wa] ~ [ikk'wa]
'mouth ' and'

b. Pre-vocalic stop: no lenition
/ip^h + i/ -> [ip^{h}i]
'leaf ' NOM

/ip + i/ -> [ibi]
'mouth ' NOM

Preservation of pre-vocalic stops in Korean is predicted by the constraint in (25), ranked above RED by J. Jun, as shown in (26):

25. Pres(mnr(-cont)): Preserve perceptual cues for noncontinuancy (p.158).

26. Partial ranking for Korean (Jun 1995: 173, adapted from his Ch. 4, (32)):
Pres(mnr(-cont)) >> RED

As stated, the constraint Pres(mnr(-cont)) requires preservation of the percept of stophood. Pre-consonantally, this constraint is satisfied in Korean during coda reduction by compensatory lengthening of the following stop closure. Although Jun (1995) did not discuss the consequence of the ranking in (26) for intervocalic stops, the hierarchy correctly predicts the non-occurrence of intervocalic lenition: Since there are no neighboring stops which may compensatorily lengthen in the intervocalic position, Pres(mnr(-cont)) would prevent reduction in this environment[32]. Consequently, the lower-ranked constraint RED is violated[33].

In contrast, Taiwanese appears to obey the reduction constraint both intervocalically and pre-consonantally[34], since non-word-initial stops may lenite in both environments. This suggests that the relative ranking of Pres(mnr(-cont)) and RED is reversed for Taiwanese, as shown in (27). The relative ranking between Pres(mnr(-cont)) and the Pres(pl) constraints is left undetermined for Taiwanese.

27. Pres(F(w.i.)) >> **RED** >> Pres(mnr(-cont)),
{Pres(pl(dor˺)) >> Pres(pl(lab˺)) >> Pres(pl(cor˺))}

---

[32] Another way of preserving the percept of stophood is by glottalization, in which case the question of whether the supralaryngeal closure is preserved is irrelevant. However, glottalization seems not to be an option in Korean in the implementation of Pres(mnr(-cont)).

[33] In a way, both constraints RED and Pres(mnr(-cont)) are satisfied in the intervocalic position, since, both the intervocalic /p^h/ and /p/ are shorter in duration compared to their non-intervocalic counterparts. It is this shortened duration that results in the voicing of the lenis /p/ in /ip i/ 'mouth-NOM' in (24b) (Sun-Ah Jun, p.c.). However, Korean intervocalic reduction differs from Taiwanese intervocalic reduction in that only in the latter case does spirantization result.

[34] In a cluster C1C2, where C1 may be reduced, the preceding word is identified as a closed syllable by Taiwanese speakers; this might be interpreted as indication of compensatory lengthening. However, non-Taiwanese speakers tend to identify the "assimilated clusters" as singleton consonants, suggesting that the native Taiwanese speakers' intuitions of consonant clusters are based on cues other than the "cluster" duration. A likely candidate is tone, since closed syllables carry different tones from open syllables, and this difference is maintained irrespective of the occurrence of place assimilation, or more precisely, coda reduction.

(28) evaluates two hypothetical concatenated strings in Taiwanese and Korean. The lowering diacritic [˕] indicates reduction. As the tableaux in (28) illustrate, for both Taiwanese and Korean, intervocalic reduction satisfies RED but violates Pres(mnr(-cont)), while intervocalic stop preservation satisfies Pres(mnr(-cont)) but violates RED. The language-specific relative ranking of the two constraints yields intervocalic lenition in Taiwanese, but not in Korean.

28 a. Taiwanese intervocalic (word-final) stops: reduction possible

aP.a	RED	Pres(mnr(-cont))
aPa	!*	
→ aP̞a		*

b. Korean intervocalic stops: reduction impossible

/ap.a/	Pres(mnr(-cont))	RED
→ apa		*
ap̞a	!*	

An additional prediction of the partial Taiwanese ranking RED >> Pres(mnr(-cont)) is that in Taiwanese, place assimilation is not obligatory, i.e., in the cluster C1C2, C2 will not necessarily lengthen as C1 reduces. This is indeed attested: To the phonetically trained ear, repetitions of the string /tsaP kʰɔ/ 'ten dollars' range from the unreduced [tsaˀpkʰɔ], to the reduced and assimilated [tsakkʰɔ], to the reduced but unassimilated (and non-overlapping) [tsaβkʰɔ], to the reduced and overlapping [tsakʰɔ].

Next, consider the observation about Taiwanese in (18e): Why is lenition detected more consistently intervocalically than pre-consonantally? Before answering this question, note first that consonantal reduction has different perceptual consequences depending on the environment: Intervocalic reduction is perceived as lenition, but pre-consonantal reduction is often heard as place assimilation. Put another way, while reduction in the pre-consonantal position may result in the perceptual deletion of place cues, the place of articulation information is more easily salvageable intervocalically, even after reduction, because of the C-to-V transition and the absence of a dominant[35], overlapping C2 (to mask the word-final consonant's place cues). Thus, the perceptual cost (i.e., information loss) of reduction is greater pre-consonantally than intervocalically. Accordingly, reduction is less frequently observed in the former environment[36]. We might ask, then, that given the presence of the C-to-V transition, why does the intervocalic, word-final consonant not preserve the stophood, which will contribute an even more salient cue in the C-to-V transition, i.e., the stop burst? Recall, however, the dominating constraint RED requires the minimization of effort, and reduction must therefore apply.

In addition to perceptual considerations, the higher frequency of lenition in V-V when compared with lenition in V-CV may be due to increased vowel-consonant-vowel co-articulatory pressure in the former environment, which would predict more target undershoot for the word-final consonant in the intervocalic position.

Let us turn now to (18f): Lenition is not perceived pre-pausally in Taiwanese. There are two potential articulatory contributors to the percept of stophood: pre-pausal glottalization and

---

[35] By a dominant C2, I mean one that influences the (acoustic) place cues present in the V-to-C1 transition.

[36] This need not be a counterexample to the Production Hypothesis, even though more effort seems to be incurred in the less salient environment: Since the crucial cues carried by the final consonants are place and nasality, and since both types of information will still be present intervocalically even with reduction, in all speech rates and styles, the effort-conserving option of intervocalic lenition is always preferred. In contrast, pre-consonantal reduction is more variable, depending on speech rate and style.

supralaryngeal stop closure. While glottalization is consistently perceived, supralaryngeal reduction is harder to detect in the presence of glottalization. Nevertheless, differences in lenition has been observed across different environments, at least for word-final labials, in an examination of the author's videotaped speech, which showed labial reduction pre-vocalically and pre-consonantally, but not utterance-finally. Therefore, one reason lenition is not perceived pre-pausally is that the oral articulation is not reduced.

With respect to final glottalization, I assume a pre-pausal glottal fortition analysis: The word-final stops are strengthened (perceptually) by glottalization, which enhances the percept of stophood at a phrasal/utterance boundary. The constraint in (29) is proposed to account for pre-pausal glottalization:

29. Glott(C#): The pre-pausal consonant must be glottalized.

We suggest that Glott(C#) may be a way of marking the end of utterances, and word-boundaries in careful speech; it may thus be considered phrasal-edge strengthening. In addition, glottalization may serve to enhance the contrast between open (CV) and closed (CVC) syllables, since the constraint in (29) will only force glottalization in pre-pausal closed syllables, and not in open syllables. (Additional cues to the syllable types include tones and vowel length.)

I propose that the constraint Glott(C#) be ranked above RED, since the reduction constraint aims to conserve articulatory effort not only by reducing supralaryngeal constrictions, but by reducing the magnitude of all articulatory gestures, including the glottal adduction gesture. The proposed ranking is shown in (30):

30. Pres(F(w.i.)), Glott(C#) >> **RED** >>
Pres(mnr(-cont)), {Pres(pl(dor̚)) >> Pres(pl(lab̚)) >> Pres(pl(cor̚))}

In (30), the relative ranking between Pres(F(w.i.)) and Glott(C#) is left undetermined, since these two constraints will never interact, and will therefore never provide evidence showing their relative ranking.

Returning to the supralaryngeal gesture of pre-pausal stops, I suggest that the demand for effort conservation is not less in the pre-pausal position. In fact, EPG studies of Taiwanese consonants currently in progress by Shu-hui Peng reveal that utterance-final /T/ and /K/ are frequently reduced in fast speech[37] (Shu-hui Peng, p.c.). The constraint ranking proposed thusfar would predict this behavior, since the constraint **RED** is not outranked by a constraint such as Pres(F(u.f.)) which demands the preservation of all features associated to the utterance-final, or pre-pausal position, including the stophood of the final consonant.

Finally, turning our attention to the observation that labial stops are less likely to reduce than non-labials in Taiwanese (18g), I would like to point out the Taiwanese-Korean asymmetry stated in (19b), repeated below:

19b. In Korean, labials and alveolars reduce to the exclusion of velars. In Taiwanese, velars and coronals are more likely to reduce than labials.

---

[37] Although my videotaped speech showed non-lenition of utterance-final labials, the observation is based only on six repetitions of /liP/ 'to enter' and /laP/ 'to pay' of one speaker under experimental condition, which should hardly be claimed as a counterexample to utterance-final lenition. Moreover, I have already suggested that word-final labials are the least likely candidates to undergo reduction, of the three places of articulation; this is a further caution in basing a non-lenition generalization on the behavior of labials.

As discussed in Section 3.3, the pattern of reduction in Korean place assimilation (i.e., coronal and labial stops may reduce to the exclusion of velar stops) is one of the possible patterns of reduction predicted by Jun's acoustically-motivated universal preservation hierarchy in (17):

17. Pres(pl(dor˺)) >> Pres(pl(lab˺)) >> Pres(pl(cor˺))

Depending on the ranking of the reduction constraint RED, a language may reduce all unreleased gestures, only unreleased labials and coronals, only unreleased coronals, or nothing. All four patterns of reduction have been attested (See Section 3.3), and Korean is an instantiation of the second possibility. Thus, (17) would suggest that, based on acoustic considerations, coronals are more likely to reduce than labials, which are in turn more likely to reduce than velars. What (17) does *not* predict is that velars would be more likely to reduce than labials, as has been observed in Taiwanese.

Is the Taiwanese reduction pattern evidence against the claim that the acoustically-motivated ranking in (17) is universal? That is, would it be appropriate to account for Taiwanese simply by reranking the constraints in (17), such that Pres(pl(lab˺)) outranks Pres(pl(dor˺)) in Taiwanese? The proposed answer here is no, since the acoustic factors encoded in (17) are based on physical constraints of the human auditory system; we do not expect these constraints to vary (in a relevant way) from speaker to speaker, nor from language to language, unless the articulations are quite different across different languages. The Taiwanese reduction pattern (velars reduce more readily than labials) should therefore not be analyzed as a counterexample to the acoustically-motivated universal preservation hierarchy in (17).

If the universal status of the preservation hierarchy in (17) is valid, how then do we account for the seemingly contradictory velar-labial reduction frequency in Taiwanese? Recall the suggested ranking in (20) for Taiwanese:

20. **RED** >> Pres(pl(dor˺)) >> Pres(pl(lab˺)) >> Pres(pl(cor˺))

Ranking RED above the entire place preservation hierarchy correctly predicts that (non-initial) stops at all places of articulation may be reduced. The relative frequency of observed reduction in Taiwanese, then, must find explanation within the nature of the constraint RED. That is, why are some places more vulnerable to reduction than others?

Recall that the constraint RED is defined by Jun (1995) as a requirement for conservation of articulatory effort. In an unreduced VCV sequence, where C is a stop consonant, the vocal tract must alternate between an open configuration for the vowels and a closed configuration for the stop. One way of conserving articulatory effort is to reduce the open-closed-open trajectory by gestural reduction in the stop consonant. The result is that the consonant is now a continuant, and is produced with relatively more open vocal tract configurations. This is the case of word-final intervocalic reduction in Taiwanese.

There are three articulatory factors we will consider here in relation to reduction. First, regarding the stop-vowel asymmetry in the openness of the vocal tract, one major contributor is the jaw. It is clear from Keating, et al's (1994) cross-language study of jaw height that stops are articulated with a higher jaw position than vowels (p. 412, inferred from Table I). One might then hypothesize that a reduced stop will be articulated with a lower jaw position (more "vowel-like") than an unreduced stop; this is the hypothesis of Beckman et al. (1992). Interpreting phonemic approximants and stops as articulatory counterparts of (non-contrastive) reduced and unreduced stops in a leniting language, Keating, et al's data seem to support this hypothesis: They found that English coronal stops [t,d] are articulated with a higher jaw position than the coronal approximant [ɹ].[38] Moreover, the jaw position is the highest for coronal stops (t), the next highest for labial stops (p), and the lowest for velar stops (k). This means that in the absence of reduction, jaw displacement during VCV is the greatest in VtV, and the least in VkV.

During reduction, one way of conserving effort is to restrict the extent of jaw displacement during VCV; the constraint *RaiseJaw embodies this restriction:

31. *RaiseJaw: Do not exceed Xmm of intersegmental jaw displacement.

(31) says that for any two adjacent segments S1 and S2, the distance traveled by the jaw from S1 to S2 must not exceed a certain distance X. Depending on the degree of reduction, X may be smaller for more reduction or greater for less reduction. Generalizing across different vowel environments, implementation of *RaiseJaw is most likely to affect VtV, and the least likely to affect VkV, since an unreduced VtV sequence is more likely to violate the maximum displacement threshold X than an unreduced, intervocalic labial or velar stop. A *RaiseJaw-driven reduction (abbreviated as "red$_{*\mathrm{RJ}}$")would then result in the reduction pattern in (32), where the symbol ">" denotes "is more likely than":

32. red$_{*\mathrm{RJ}}$(cor) > red$_{*\mathrm{RJ}}$(lab) > red$_{*\mathrm{RJ}}$(dor)

Note that (32) is *not* a constraint hierarchy, but simply a descriptive statement predicted by the constraint *RaiseJaw. (32) indicates that *RaiseJaw is most likely to result in reduction in coronals, then in labials, and finally in velars. The coronal-non-coronal asymmetry in (32) is attested in English, in which coronal stops in weak positions undergo flapping[39], but non-coronals never do. However, (32) does not make the correct predictions for Taiwanese: It is true that word-final coronals are more likely to undergo lenition than labials. However, the extent of perceived reduction is less in coronals than in velars, since intervocalic coronal finals are perceived as flaps while velars in the same environment are perceived as approximants. Moreover, (32) predicts that labials should reduce more readily than velars; the opposite is observed for Taiwanese. We conclude, then, that while the constraint *RaiseJaw and the resulting reduction hierarchy in (32) may be active in Taiwanese, it is not the only active constraint underlying Taiwanese reduction. We now turn to a second articulatory component in reduction.

First, we assume that for every speech rate and style, there is an articulator-specific "comfortable" velocity X during articulator displacement, and that acceleration beyond this velocity X is disfavored[40]. This proposal is captured by the *Accel constraint in (33):

33. *Accel: Do not increase gestural stiffness beyond some articulator-specific
    peak velocity X for the given speech style.

In (33), stiffness refers to the ratio of the articulator's peak velocity to its peak displacement; "less stiffness results in slower movement towards the target" (Hawkins, 1992: 16).

Next, we assume that during vowel-consonant-vowel co-articulation, the demand for articulator acceleration is increased, since lack of acceleration may result in target undershoot. We propose that in Taiwanese, *Accel outranks the entire Pres(pl) hierarchy in word-final positions,

---

[38] Similarly, Swedish coronal stops [t,d] have higher jaw positions than the apical trill [r].

[39] I speculate that flaps are produced with a lower jaw position than alveolar stops. Furthermore, the low position of the jaw is one contributing factor to the flap's short duration, since, more effort would be required in tongue tip raising (from a low jaw position) to maintain "long" closure during a flap.

[40] Implicit in the *RaiseJaw constraint is also the avoidance of jaw movement acceleration. *RaiseJaw is thus a particular instantiation of *Accel.

where lenition is attested, and that the ranking is reversed[41] in word-initial positions, where stop preservation is attested.

What is the pattern of propensity to stop reduction predicted by *Accel? Recall our assumption that acceleration may be needed to avoid target undershoot to the extent that the neighboring gestures involve the same articulators (but potentially spatially-opposing targets). Therefore, the articulator that is shared in both vowel and stop formation, e.g., the tongue dorsum, will be the most likely to result in target undershoot due to *Accel, since the required configurations for the neighboring vowels prevent a possible (temporally-)lengthened, but unaccelerated trajectory of the shared articulator for the consonant. In this view, then, the reduction pattern in (34) is predicted by *Accel:

34. $\text{red}_{*\text{Accel}}(\text{dor}) > \text{red}_{*\text{Accel}}(\text{cor}) > \text{red}_{*\text{Accel}}(\text{lab})$

As is the case in (32), (34) is *not* a constraint hierarchy. (34) indicates that *Accel-driven reduction is the most likely to result in velar lenition. This is so since the tongue dorsum is the primary active articulator in vowel production. The next likely candidate of reduction is the tongue tip, since the tongue tip is still somewhat dependent on the tongue body (crucial for vowels) in articulation. (Recall our observation during the Po experiment that front-back vowel co-articulation may result in retraction of the medial [t], a testimony to the dependence between coronal consonantal articulation and dorsal vowel articulation.) Finally, the lips are the most independent from the primary "vowel articulator," since they are not connected in any direct sense. We would then predict least *Accel-driven reduction in the labial consonants. (34) nicely captures the labial-non-labial asymmetry in Taiwanese (18g). Furthermore, it predicts more velar reduction than coronal reduction, also assumed to be the case for Taiwanese. The constraint *Accel may thus be considered an active, if not dominating, reduction constraint for Taiwanese.

Finally, an additional, perhaps language-specific, factor which may contribute to the less likely reduction of Taiwanese labials is the extent of upper lip involvement in my production of Taiwanese and English /p/[42]. During a videotaped segment of my own speech, variable upper-lip lowering has been observed during Taiwanese /p/. The upper-lip lowering seems to be optional, since its occurrence cannot be generalized to, nor predicted by, the prosodic environment. In contrast, no upper lip lowering was observed during my production of the English [p]. This suggests that in Taiwanese, there is an additional potential active articulator in the production of labials. Both velars and coronals involve the raising of two active articulators (jaw plus tongue dorsum or tongue tip) toward a passive articulator, the roof of the mouth. In contrast, labials may behave either as having two active articulators (jaw plus lower lip) in the absence of upper lip lowering, or as having three compensating active articulators (jaw, lower lip, plus upper lip) when upper lip lowering occurs. Given the same amount of raising activity, the gesture which consists of an additional compensating, lowering activity by the target will be most likely to achieve closure, since less raising will be required by the other active articulators. Thus, we have observed during the oral pressure experiment that pre-consonantal labials alternate between reduced and unreduced articulations while pre-consonantal non-labials almost always reduce.

In sum, we have considered three possible articulatory factors in (effort-conserving) reduction for Taiwanese. These factors no doubt interact with each other. For instance, less jaw raising in the coronals would require more (compensation in the form of "extra") tongue tip raising if stop closure is to be achieved. Since we do not have experimental data to assess the relative

---

[41] The "reversal" of the relative ranking between *Accel and Pres in the word-initial position is encoded in the ranking Pres(F(w.i.)) >> *Accel. Note that the relative ranking between *Accel and Pres(pl) actually remains the same in both word-initial and word-final positions.

[42] The author is a native speaker of Taiwanese, and a near-native speaker of American English.

significance of the proposed articulatory constraints, we will simply hypothesize a *descriptive* constraint hierarchy based on the observation in Taiwanese, i.e., the constraint hierarchy in (35):

35. RED(dor) >> RED(cor) >> RED(lab)

The ranking in (35) is tentatively assumed to be the result of language-specific articulatory interactions in Taiwanese (i.e., the net effect of *RaiseJaw, *Accel, and possibly a constraint which requires upper lip compensation during the articulation of labial consonants). (35) predicts that velars will be the most likely candidate of reduction, followed by coronals, followed by labials.

The entire range of the reduction hierarchy in (35) is ranked above the Pres(pl) hierarchy for Taiwanese, as shown in (36)[43]:

36. Glott(C#), Pres(F(w.i.)) >>
RED(dor) >> RED(cor) >> RED(lab) >>
{Pres(pl(dor˺)) >> Pres(pl(lab˺)) >> Pres(pl(cor˺))}, Pres(mnr(-cont))

Since the reduction hierarchy outranks the entire preservation hierarchy, the propensity to reduction will be determined by the reduction hierarchy, crucially, when the reduction pattern predicted by the two hierarchies are in conflict. A possibly language-specific reduction hierarchy can also be assumed for Korean. However, the outranking Pres(pl(dor˺)) constraint may conflict with, and thus, dominate the determination of the reduction pattern.

Thus, two major types of reduction are exemplified by Taiwanese and Korean. In Korean-type reduction, the reduction pattern is restricted to a proper subset of the places of articulation, since only for a subset of places is the relevant preservation constraint outranked by the reduction constraint. In contrast, in Taiwanese-type reduction, lenition is not restricted to particular places of articulation, since for every place of articulation, the reduction constraint outranks the corresponding preservation constraint. Of course, both types of reduction are the consequence of a compromise between articulatory and perceptual demands. They differ simply in the ways the compromises are reached.

5.3 Summary

The constraint hierarchy in (36) has been proposed to account for the observations about Taiwanese summarized in Section 5.1, restated below:

36. Glott(C#), Pres(F(w.i.)) >>
RED(dor) >> RED(cor) >> RED(lab) >>
{Pres(pl(dor˺)) >> Pres(pl(lab˺)) >> Pres(pl(cor˺))}, Pres(mnr(-cont))

18 a. Word-final stops never have audible bursts, a fact which can be attributed to either overlap by a following consonant or to glottalization in the pre-pausal position.
b. Word-final stops at all places of articulation reduce intervocalically.
c. Word-final stops at all places of articulation may reduce pre-consonantally.
d. Word-initial stops never undergo lenition.
e. Lenition is detected more consistently in the intervocalic position than in the pre-consonantal position.
f. Lenition is not perceived pre-pausally.
g. Labial stops are less likely to reduce than non-labials.

---

[43] A more explanatory, but less descriptively explicit, account would substitute the reduction hierarchy in (36) by the (unranked?) constraints *RaiseJaw and *Accel.

To recapitulate, (non-utterance-final) reduction of word final (unreleased) stops (18a,b,c) is accounted for by the partial ranking RED(dor) >> RED(cor) >> RED(lab) >> Pres(pl(dor˺)) >> Pres(pl(lab˺)) >> Pres(pl(cor˺)). The non-lenition of word-initial stops (18d) is predicted by the ranking Pres(F(w.i.)) >> RED(dor) >> RED(cor) >> RED(lab). The observation that lenition is more prevalent intervocalically than pre-consonantally (18e) is argued to be due to the relative ease of contrast preservation in the two environments when gestural reduction occurs; the presence of the C-to-V transition in V-V, and its absence in V-CV, translate into a "smaller cost," in terms of place-of-articulation information loss, of reduction in the former. Next, the "non-lenition" of pre-pausal stops (18f) is the consequence of the partial hierarchy Glott(C#) >> RED(dor) >> RED(cor) >> RED(lab); glottalization ensures the percept of stophood even in the presence of supralaryngeal reduction (predicted by the partial hierarchy RED(dor) >> RED(cor) >> RED(lab) >> Pres(pl(dor˺)) >> Pres(pl(lab˺)) >> Pres(pl(cor˺)).). The higher tendency of non-labial reduction (18g) is captured by the descriptive hierarchy RED(dor) >> RED(cor) >> RED(lab), posited based on the observations about Taiwanese, and hypothesized to be the net effect of *RaiseJaw and *Accel.

5.4 Conclusion

In this section, we have provided a phonetically-based analysis of reduction in Taiwanese. Adopting the framework of the Optimality Theory, the reduction and non-reduction patterns in Taiwanese are shown to be the result of a particular reconciliation of the potentially conflicting goals of ease of articulation and ease of perception, embodied in the constraints proposed here. More specifically, we have argued that consonant reduction is a general effort-conserving phenomenon in Taiwanese, prohibited only in the word-initial position where preservation of informative acoustic cues via the medium of unreduced gestures is crucial to lexical access. After all, we speak to be heard to be understood, but efficiently so!

Appendix A: List of Taiwanese phrases for the oral pressure experiment (Section 4)

In the English glosses below, '*POSS*' stands for a Possesive marker, '*Cl.*' stands for Classifier, '*REL*' stands for Relativizer, '*DIM*' stands for Diminutive suffix and '*FAM*' stands for Familiar suffix.

1. li e papa — 'your dad'
you POSS dad

2. li e mama — 'your mom'
you POSS mom

3. gɔ tsaP e laŋ — 'fifty people'
five ten Cl. people

4. li lam e laŋ — 'the people you are embracing'
you hug REL people

5. i le tɕiP ku a — 'he is collecting beetles'
he at collect beetle DIM

6. i e kim ku a — 'his beetles'
he POSS gold beetle DIM

7. i le tɕiP tu a — 'he is collecting cabinets'
he at collect cabinet FAM

8. i e kim tu a — 'his golden cabinet'
he POSS gold cabinet FAM

9. i le kʰiK pu a — 'he is sculpting gourds'
he at carve gourd DIM

10. i le tɕiŋ pu a — 'he is growing gourds'
he at grow gourd DIM

11. i le tɕʰiT pu a — 'he is wiping the gourd'
he at wipe gourd DIM

12. i ɕiaŋɕin puttsɔ[44] — 'he believes in Buddha'
he believe Buddha

---

[44] In the dialects of Speakers f2, m3, and m4, the word for 'Buddha' is pronounced [huttsɔ], with an initial [h] rather than [p]. These speakers produced the initial [p] as shown in (12) only upon my request.

Appendix B: Summary of Po results by speaker

Each table below shows the resulting Po trace for CC for one speaker using the notations described in Section 4.4. The sequences analyzed are shown in the left-most column; the top row indicates the repetition (or the medial six) under consideration. The speaker identity is listed on the top left corner: m1 for Male Speaker 1, f1 for Female Speaker 1, etc.

m1	R1	R2	R3	R4	R5	R6
iPku	PN	PN	PN	PN	PN	PN
imku	PN	PN	PN	PZ	PN	PN
iPtu	PZ	PZ	PZ	PZ	PZ	PN
imtu	PZ	PN	PN	PZ	PZ	PZ
iKpu	AP	AP	AP	AP	AP	AP
iŋpu	AP	AP	AP	AP	AP	AP
iTpu	AP	AP	AP	AP	AP	AP
inpu	AP	AP	AP	AP	AP	AP

m2	R1	R2	R3	R4	R5	R6
iPku	PZ	PZ	PZ	PZ	PN	PZ
imku	PN	PN	PN	PN	PN	AZ
iPtu	PZ	PZ	PZ	PZ	PZ	PZ
imtu	AZ	AZ	AZ	AZ	AZ	AZ
iKpu	AP*	AP	AP*	AP	AP	AP
iŋpu	AP	AP	AP	AP	AP	AP
iTpu	AP	AP	AP	AP	AP	AP
inpu	AP	ZP	ZP	ZP	ZP	ZP

m3	R1	R2	R3	R4	R5	R6
iPku	PN	PZ	PZ	PN	AZ	PN
imku	PN	PN	PN	PN	PN	PN
iPtu	AZ	PZ	PZ	PZ	PZ	PZ
imtu	PN	PN	PN	PN	PN	PN
iKpu	AP	AP	AP	AP	AP	AP
iŋpu	AP	AP	AP	AP	AP	AP
iTpu	AP	AP	AP	AP	AP	AP
inpu	AP	AP	AP	AP	AP	AP

---

* Two tokens of iKpu showed positive Po throughout the Kp duration, with a brief return to zero Po during the middle of Kp. I have analyzed these two tokens as instantiations of AP, interpreting the medial zero *not* to be an interruption during two distinct, sequenced [p]s, but as the result of the pressure tube opening coming in contact with, and thus blocked by the teeth, resulting in zero pressure. This interpretation is plausible because Speaker m2 expressed concern earlier during the recording session about the tube coming into contact with his teeth. Furthermore, similar pressure traces during an intervocalic [p] resulted for the same speaker from an independent token not recorded for the purpose of this experiment, where we have no reason to posit two distinct [p] gestures. (See Appendix C)

m4	R1	R2	R3	R4	R5	R6
iPku	PZ	PZ	PZ	PZ	PZ	AZ
imku	PZ	PN	PN	PN	AZ	AZ
iPtu	PN	PZ	PZ	PZ	PZ	PZ
imtu	PN	PN	PN	PZ	AZ	PZ
iKpu	AP	AP	AP	AP	AP	AP
iŋpu	AP	AP	ZP	AP	AP	AP
iTpu	AP	AP	AP	AP	AP	AP
inpu	AP	AP	AP	AP	AP	AP

f1	R1	R2	R3	R4	R5	R6
iPku	PN	PN	PN	PN	PN	AZ
imku	PN	PN	PN	PN	PN	PN
iPtu	PN	PN	PZ	PN	PZ	AZ
imtu	PN	PZ	PN	PN	PN	PN
iKpu	AP	AP	AP	AP	AP	AP
iŋpu	NP	AP	NP	NP	NP	NP
iTpu	AP	AP	AP	AP	AP	AP
inpu	AP	AP	AP	AP	AP	AP

f2	R1	R2	R3	R4	R5	R6
iPku	AZ	AZ	AZ	AZ	AZ	AZ
imku	ZN	AZ	ZN	PN	ZN	ZN
iPtu	AZ	AZ	AZ	AZ	AZ	AZ
imtu	PZ	PZ	PZ	AZ	AZ	AZ
iKpu	AP	AP	AP	AP	AP	AP
iŋpu	AP	AP	AP	AP	AP	AP
iTpu	AP	AP	AP	AP	AP	AP
inpu	AP	AP	AP	AP	AP	AP

Appendix C: Examples of the Medial-Zero Po pattern by Speaker m2

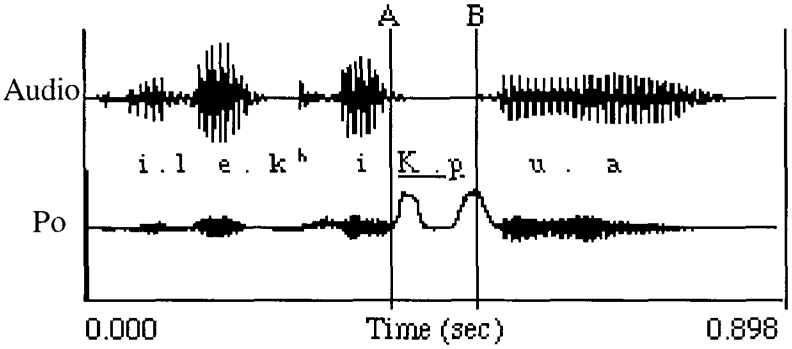


Figure C1: Example of Speaker m2's production of iKpu showing a medial zero in Po during the /Kp/ duration. The top waveform is the audio; the bottom is the Po trace. The vertical axis measures pressure while the horizontal axis measures the time. I have categorized the Po pattern between points A and B as an instantiation of AP, interpreting the medial zero to be the result of pressure tube blockage by the teeth.

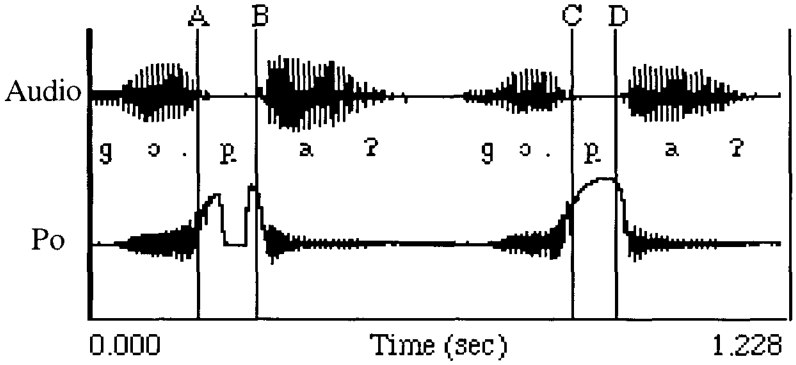


Figure C2: Two examples of intervocalic, initial [p] by the same speaker for the utterance [gɔ paʔ] 'five hundred'. The example on the left shows the same pattern as in Fig. C1: positive pressure during the C(C) duration with a medial zero (the interval between A and B). The example on the right shows no medial zero (the interval between C and D): There was no pressure tube blockage during the production of this token.

References

Barry, Martin (1992). Palatalisation, Assimilation and Gestural Weakening in Connected Speech. *Speech Communication* 11: 393-400.

Beckman, Mary, Jan Edwards and Janet Fletcher (1992). Prosodic Structure and Tempo in a Sonority Model of Articulatory Dynamics. In G. J. Docherty & D.R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody.* Cambridge: Cambridge University Press, 68-86.

Browman, Catherine P. and Louis M. Goldstein (1986). Towards an Articulatory Phonology. *Phonology Yearbook* 3: 219-252.

Browman, Catherine P. and Louis M. Goldstein (1989). Articulatory Gestures as Phonological Units. *Phonology* 6: 201-251.

Browman, Catherine P. and Louis M. Goldstein (1990). Tiers in Articulatory Phonology. In John Kingston and Mary E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech.* Cambridge: Cambridge University Press, 341-376.

Browman, Catherine P. and Louis M. Goldstein (1992). Articulatory Phonology: An Overview. *Phonetica* 49: 155-180.

Brown, Roger and D. McNeill (1966). The Tip-of-the-Tongue Phenomenon. *Journal of Verbal Learning and Verbal Behavior* 5: 325-337.

Byrd, Dani (1992). Perception of Assimilation in Consonant Clusters: A Gestural Model. *Phonetica* 49: 1-24.

Byrd, Dani (1994). *Articulatory Timing in English Consonant Sequences.* PhD dissertation, UCLA, distributed as *UCLA Working Papers in Phonetics* 86.

Dorais, Louis-Jacques (1986). Inuktitut surface Phonology: A Trans-dialectal Survey. *International Journal of American Linguistics* 52.1: 20-53.

Fougeron, Cécile and Patricia A. Keating (1995). Demarcating Prosodic Groups with Articulation. Paper presented at the Acoustical Society of America in Washington DC.

Hawkins, Sarah (1992). An Introduction to Task Dynamics. In G. J. Docherty & D.R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody.* Cambridge: Cambridge University Press, 9-25.

Jun, Jongho (1995). *Perceptual and Articulatory Factors in Place Assimilation: An Optimality Theoretic Approach.* PhD dissertation, UCLA.

Keating, Patricia A., Björn Lindblom, James Lubker and Jody Kreiman (1994). Variability in Jaw Height for Segments in English and Swedish VCVs. *Journal of Phonetics 22*: 407-422.

Kohler, K. J. (1991). The Phonetics/Phonology Issue in the Study of Articulatory Reduction. *Phonetica* 48: 180-192.

Lin, Yen-Hwei (1989). *Autosegmental Treatment of Segmental Processes in Chinese Phonology*, PhD dissertation, University of Texas at Austin.

Lindblom, Björn and Ian Maddieson (1988). Phonetic Universals in Consonant Systems. In L. Hyman & C. Li (eds.), *Language, Speech, and Mind.* London: Routledge, 62-78.

Lindblom, Björn (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modelling.* Dordrecht: Kluwer Academic Publishers, 403-439.

Maddieson, Ian (1984). *Patterns of Sound.* Cambridge: Cambridge University Press.

Martinet, André (1952). Function, Structure, and Sound Change. *Word* 8: 1-32.

McCarthy, John & Alan Prince (1993). *Prosodic Morphology I: Constraint Interaction and Satisfaction.* Ms. University of Massachusetts, Amherst, Rutgers University.

Mohanan, K. P. (1993). Fields of Attraction in Phonology. In J. Goldsmith (ed.), *The Last Phonological Rule: Reflections on Constraints and Derivations.* Chicago: University of Chicago Press, 61-116.

Nolan, Francis (1992). The Descriptive Role of Segments: Evidence from Assimilation. In G. J. Docherty & D.R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody.* Cambridge: Cambridge University Press, 261-280.

Peng, Shu-hui (In progress). *Phonetic Implementation and Perception of Segmental Coarticulation and Tone Sandhi.* PhD dissertation, Ohio State University.

Prince, Alan & Paul Smolensky (1993). *Optimality Theory: Constraint Interaction in Generative Grammar.* Ms. Rutgers University, University of Colorado, Boulder.

Silverman, Daniel & Jongho Jun (1994). Aerodynamic Evidence for Articulatory Overlap in Korean. *Phonetica* 51: 210-220.

Steriade, Donca (1993). Neutralization and the Expression of Contrast. Paper presented at NELS 24.

Tung (1957). Xiamen Phonology. *Bulletin of the Institute of History and Philology XXIX, Part I: Studies Presented to Yuen Ren Chao on His Sixty-fifth Birthday.* Taipei: Academia Sinica, 231-253.

Zhang, Z. X. (1983). *Taiwan Minnan Fangyan Jilue* (Notes on Southern Min Dialects of Taiwan). Fuzhou: Fujian People's Press.

Zsiga, Elizabeth (1992). Acoustic Evidence for Gestural Overlap in Consonant Sequences. *Haskins Laboratories Report on Speech Research* 1992, SR-111/112: 1-20.

The Phonology-Phonetics Interface

Patricia A. Keating

1. Introduction

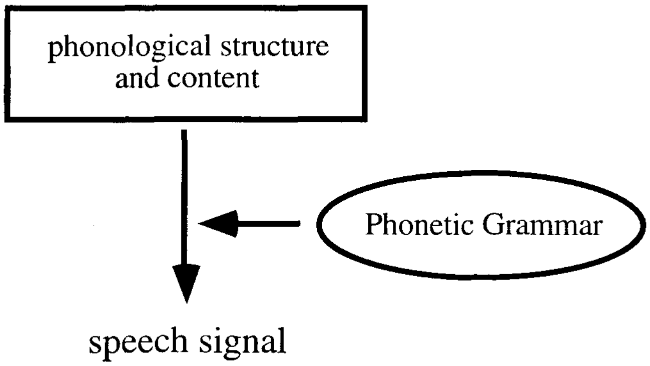
The term "Phonology-Phonetics Interface" can refer to many things. Some topics that come under this rubric which I will not be able to discuss here are listed in (1). A number of recent UCLA dissertations which treat topics of these kinds are included among the references.

(1) Some topics that traditionally fall under "The Phonology-Phonetics Interface":

- motivating the elements of phonological description
  - features, segments, syllables, feet, ...
  - sonority, stress, ...
- explaining typological findings
  - segments, contrasts, inventories
  - synchronic, diachronic processes (after Keating 1988a)

Instead, I will talk about a central question in linguistic phonetics, namely, how relatively abstract phonological structure and content is physically realized — how articulators are moved, how sound is acoustically and therefore auditorily shaped, so that the linguistic structure is made available to a listener. The answer to this question has been conceived of as a component of grammar: a phonetic component that relates the output of the phonology and observable aspects of speech events. This relation has been conceived of as a derivational one, and cast in terms of phonetic rules. These rules are sometimes called rules of ***phonetic implementation.***

(2) Phonetic implementation as the mapping of phonological structure into phonetic output (Following Pierrehumbert 1980).



This conception assumes in turn an answer to another question: Where is the boundary between phonology and phonetics? The basic answer underlying most work in this framework is that *phonology* deals in **discrete symbolic elements**, while *phonetics* deals in **numbers** (on continuous dimensions).

Because phonology deals in symbolic representations, it incorporates certain idealizations from the physical speech signal, including chunking speech in temporal and qualitative dimensions. These chunks can then be represented by symbolic elements, e.g. feature values. Phonology also then confines its operations to operations on and between these symbolic elements. So phonological operations are *categorical*, meaning that category labels are affected.

Because phonetics deals in numbers, it relates these phonological idealizations to speech. It undoes chunking in time to deal in real (or at least notional) time; it undoes chunking in qualities to deal in continuous dimensions. Gradual changes over time along quality dimensions are said to be ***gradient***.

These different properties associated phonological and phonetic representations and processes are summarized in (3). Later I will give examples of phonetic effects that should help make this distinction clearer.

(3) *Phonology*
symbolic representations
allow idealizations of
    temporal chunking (segmentation)
    qualitative categorization (labels)
    timelessness
rules manipulate
    features and feature values
    associations
thus phonological rules can be category changing, produce static changes over whole segments; can be lexical/cyclic

Phonetics
physical representations
continuous in time and space
internal temporal structure
    allows overlap
quantitative values on multiple independent dimensions
rules interpret feature values in time and space, can be gradient
(after Keating 1988b, Cohn 1990)

2. Some models

Autosegmental/CV Phonology was an important development from the point of view of relating phonology and phonetics. Goldsmith (1976)'s break with strict segmentation was crucial because speech characteristics are not always chunked into neat segments. Strictly segmental SPE phonological representations had made it harder to bridge this gap conceptually.

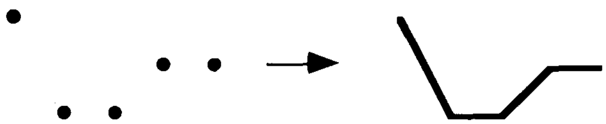
The traditional class of models of this relationship (which include many speech synthesizers) is summarized under (4) and can be described as providing ***targets***, and ***interpolations*** between targets. Features (or segments) specify "targets" that the human articulators or a computer synthesizer aim at, moving, or "interpolating", from target to target. These can be stated in articulatory or acoustic or auditory domains.

(4) **Targets and Interpolation**

Step 1: Relate a feature to one or more parameters in the domain (e.g. articulatory)
e.g. [Nasal] --> velic opening
e.g. [Voice] --> glottal configuration **AND** other parameters that facilitate or prevent vocal cord vibration

Step 2: Interpret value of feature as value(s) along parameter(s)
e.g. [+nasal] --> some amount of velic opening over some time interval
e.g. [+voice] --> vocal cords appropriately positioned and tensed for vibration, (for a stop) oral tract walls lax

Step 3: Connect up the targets according to some function



Result: A plan for each articulator to follow, or a predicted acoustic output .

The first two steps result in a sequence of targets along each dimension. The third step connects the targets in continuous space, resulting in a plan for an articulator to follow, or a predicted acoustic contour. The important point is that conceptually, the targets are what realize the feature values, and they are prior to the interpolation function. Two instances of such models are the intonational model of Pierrehumbert 1980 (and subsequent work) and the segmental models in my own work (and subsequent).

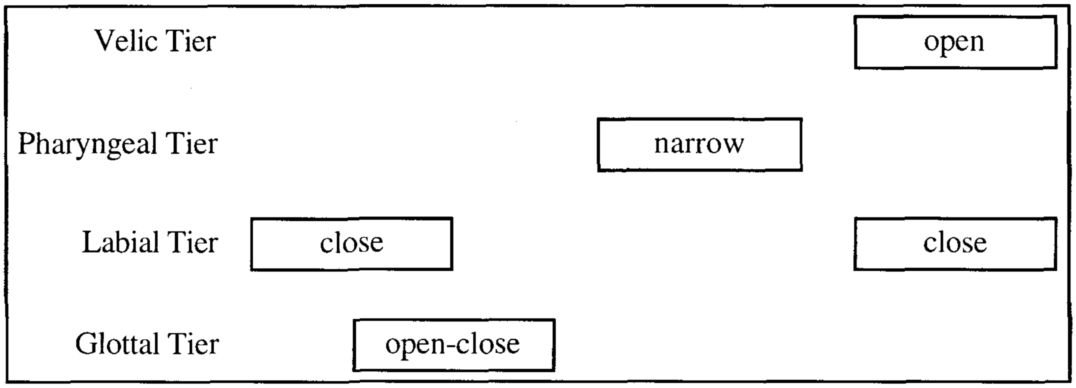
A similar way of making this distinction has been proposed by Hertz (e.g. Hertz 1991; Clements and Hertz, ms). The acoustic speech stream is segmented into *phones*, which are quasi-steady-state targets, and *transitions* which are the interpolated intervals between phones.

Targets on each physical dimension are completely autosegmentalized. For example, the different articulations for a single segment can begin and end at different times. This relatively free alignment results in overlaps, where part of an articulation from one segment occurs at the same time as an articulation from a different segment. This overlap is traditionally called co-articulation. It needs to be stressed that this is something we control when we speak — articulations can be more or less overlapped, and languages differ in this respect. Target models in general have tended to focus more on the individual dimensions than on their inter-alignment.

In contrast, the Articulatory Phonology model of Browman & Goldstein (e.g. 1992) focuses more on these inter-alignments, or ***phasings***. This theory also conceives of events on the individual dimensions differently. In a Targets and Interpolation model, we think of the targets as primary and the movements as at best secondary. In contrast, we could think of the movements towards and away from the targets, and their durations, as potentially equally important. In Articulatory Phonology, the phonological primitives are not features, or segments that they can belong to, but abstract articulatory gestures. Within words, gestures are specified for their durations so that the whole time-course is primary. When targets are represented, they are represented as having notional durations. The example frequently given by Browman &

Goldstein is schematized in (5), which shows the targets of articulatory gestures in the initial (presumably lexical) representation of the word "palm".

(5) Schematic gestural score for “palm” [pʰɑm] (after Browman & Goldstein 1990)

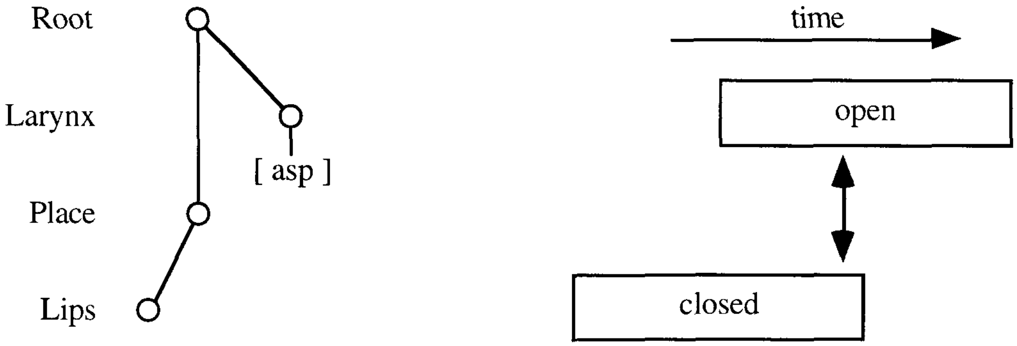


In this schematic, time runs from left to right and articulatory tiers run from top to bottom. The first event is a labial closure (made by the lips); this is closely followed by glottal spreading, indicated as a sequence of opening and closing of the vocal cords — i.e. assuming the velum has been closed, the two together make an aspirated [pʰ] Next comes a constriction in the pharynx (for a low back vowel). The last events, which are indicated as simultaneous, are another labial closure and a velic opening — i.e. nasal [m]. Note that there is no segmentation *per se*. The [ɑ] looks like a segment because it has only one gesture specified; and the [m] looks like a segment because its two gestures are simultaneous. But the [pʰ] looks like a segment only insofar as its two gestures are associated; they are not simultaneous. Note also that this timing relation says that the aspiration of this [pʰ] is lexically specified; it is not derived allophonically.

Articulatory Phonology aims for a closer relation between phonetics and phonology by making phonological representations more like phonetic ones, using the same primitives (the gestures) and putting a fair amount of predictable information into lexical representations. Nonetheless, even this theory needs phonetic implementation just as Targets and Interpolations models do, in the form of postlexical adjustments to the pre-specified gestures and alignments, and a decision mechanism for allocating gestures to articulators (called Task Dynamics). For example, the two gestures of the [m] in (5) are shown as simultaneous. Yet this [m] is in coda position, where a velic opening gesture comes before a labial gesture (Krakow 1989). Therefore presumably some realignment occurs between these gestures from what is represented here.

Compare this approach with Zsiga 1993: she treats Articulatory Phonology not as a combination of phonology and phonetics, but as a model of the phonetic component that is an alternative to Targets & Interpolation models. Zsiga maintains the distinction between categorical phonological and gradient phonetic representations and rules that are central to the Targets and Interpolation framework, but transforms phonological feature values into articulatory gestures, in a kind of Articulatory Phonetics. An example from her dissertation is shown in (6). This is an appealing approach but there are some relations between features and gestures that are more complex than this model is designed to handle, because it assumes that every feature maps to one gesture.

(6) Time-aligned gestural score (right) as a phonetic elaboration of timeless phonological features (left; after Zsiga 1993)



Another model which takes something like traditional phonological features as the input to phonetic implementation, and which takes dynamic gestures rather than targets to be the phonetic primitives, is Fujimura's C/D model (e.g. Fujimura 1994). This model is more concerned with questions of prosodically-conditioned and speaker-conditioned timing and scaling, than with the distinction between target models and non-target models.

Rather than go into the differences among these models, I would like to make a more general point more germane to the interface between phonetics and phonology. A result of work under both kinds of phonetic models is the following: what has been taken to be "phonological" assimilation can be phonetic, by which I mean gradient rather than categorical. Many examples are available; here I will discuss four.

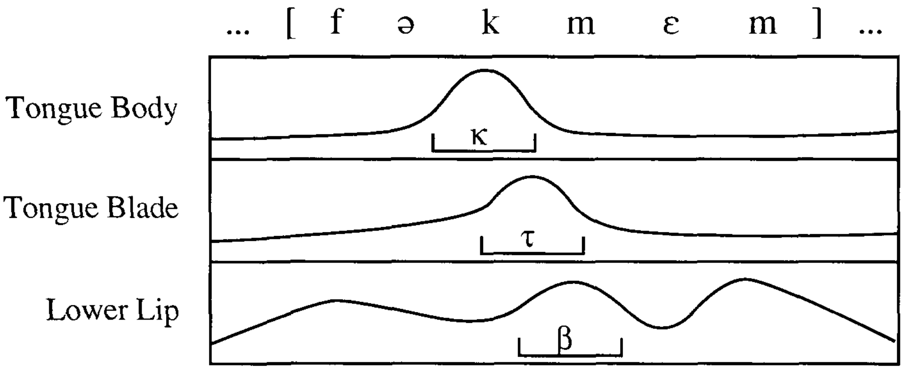
3. Phonetic assimilation

3.1. When different articulations compete for some interval of time

3.1.1. Overlap in English

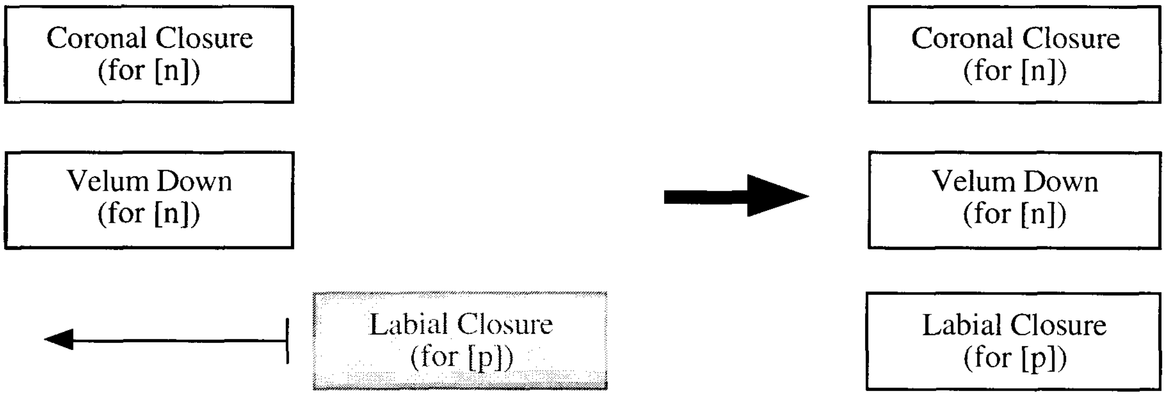
In English, and in many languages, consonants in clusters overlap in time, and the faster we speak the more this is so (discussed at length by Byrd 1994). Browman and Goldstein used this fact to explain the optional deletion of coronal stops between two other consonants. Their now-famous example was the phrase "perfect_memory" in which no [t] is heard. Usually if no /t/ is heard it would be described as having been deleted. It might be expected that as a result of deletion any features of the [t] that might provide a phonetic target are absent. The phonetic data presented by Browman and Goldstein bearing on this phonological deletion analysis are schematized in (7). These are records of the movements of small pellets attached to articulators (listed on the left of picture) during a token of "perfect memory" transcribed without [t]. (Browman and Goldstein use Greek letters for gestures, here kappa, tau, beta for three active aticulators.) This picture shows that there is a clear tongue blade articulations for a [t], as well as for the [k] and [m] also. The key thing is not the presence or absence of a [t] articulation, but its timing relative to the other two stop articulations. The [t] starts just after [k] and ends during [m]. The [t] is completely hidden because of the temporal overlap of the different articulations. It is deleted only in the sense that it is inaudible. Browman and Goldstein suggested that this is the typical situation in deletion.

(7) Coronal place assimilation (after Browman & Goldstein 1990)



They then suggested that the same is true of some place assimilations, like "seven plus" where the /n/ appears as [m]. A schematic of this is shown in (8). The /p/'s labial articulation overlaps with the /n/'s coronal articulation, so that even if the /n/ is not reduced, its lingual articulation is not heard. On the other hand, the nasalization for the /n/, which now also overlaps with the /p/, is not rendered inaudible by this overlap, but instead makes part of the /p/ an [m].

(8) Schematic of n --> m / __ p, as in "seven plus"

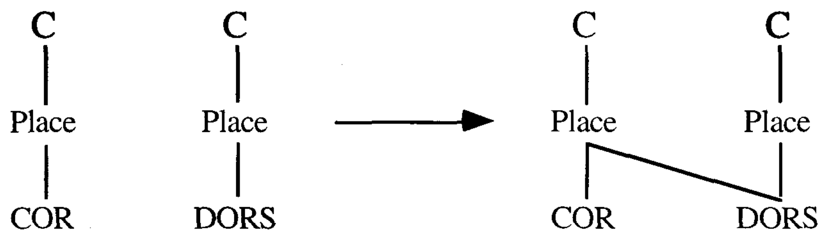


Browman & Goldstein did not actually observe this timing pattern, and since then various people, most recently J. Jun at UCLA (J. Jun 1995), have shown that this kind of assimilation, in which both articulations are intact, is probably pretty rare. Nonetheless, the point Browman & Goldstein made still holds in principle: that the same kind of event, overlap, could lead to apparent deletion in one case and apparent assimilation in another.

Hayes (1992) has discussed how such cases may be treated as phonological spreading: the Place node of the last consonant spreads onto the preceding consonant without delinking any of the preceding consonant's own features, as reproduced in (9). This is phonological overlap, as for a double-articulation or "complex segment".

(9) Phonological overlap (after Hayes 1992):

"late calls": /tk/ --> $\left[\left\{\begin{smallmatrix}k\\t\end{smallmatrix}\right\} k\right]$



So the point is not that articulatory overlap cannot be represented phonologically. The point is Browman and Goldstein's suggestion that the degree of overlap depends on rate or casualness of speech, so that the resulting assimilation occurs more-or-less, not all-or-none. It's not simply that the spreading is optional but more likely to occur in fast speech; instead the overlap usually does occur, but to different degrees in different situations, only the most extreme of which result in auditory deletion or assimilation. Phonological spreading is not generally taken to operate gradiently in this way, whereas phonetic implementation must.

3.1.2. Korean Lenis Stop Voicing

In Korean, lenis stops are voiced when they are between vowels, as could be crudely described by the rule in (10). S. Jun's example sentence includes two examples, a word-medial and a cross-word case.

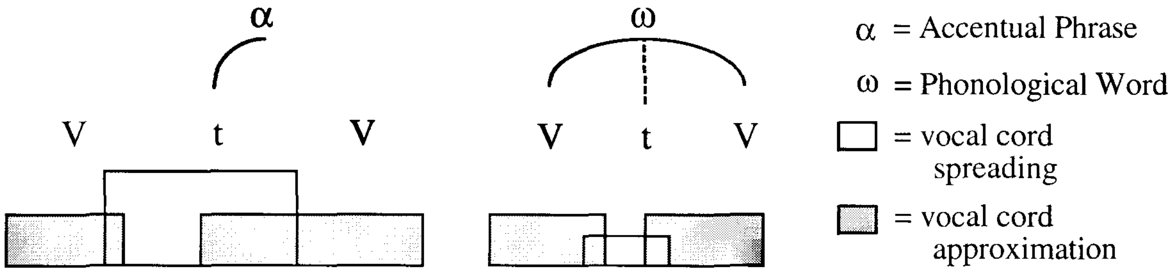
(10) Lenis stop voicing (example from S. Jun 1993) as a phonological rule

Example: [jəlmeg̲a#d̲aɾannja], *Was the FRUIT sweet?*

Phonological Rule: [-spread, -constricted] --> [+voice] / V __ V

The rule as written is assimilatory, suggesting that the glottal configuration normally specified by a lenis stop is given up in favor of the glottal configuration of the surrounding sonorants, so that there is one target through all three segments. Instead, Jun shows that the voicing is a function of the relative phonetic duration of the consonant. Short lenis consonants next to long voiced segments voice. Her interpretation of this is as follows: a lenis stop which is voiceless is known to involve some opening of the vocal cords — not as much as for an aspirated stop, but enough to prevent voicing (sometimes enough to make the vowel breathy). This opening requires some minimum amount of time; if there is too little time, then the opening won't be enough to prevent voicing, and the consonant will be voiced. When will there be too little time? When the consonant itself is short, because of its prosodic position, and when the segments on either side of the stop involve glottal approximation, not opening. The glottal targets cannot all be achieved in the available time, and they compete. The voiced segments then win out over the lenis consonant if that consonant is short. Jun's schematic of this situation, in terms of articulatory gestures, is reproduced in (11).

(11) Lenis stop voicing as a phonetic effect



There are three glottal gestures (for V, C, V), in one case initial in the Accentual Phrase, in the other case not. The glottal gesture of the consonant is longer in Accentual Phrase initial position than it is when internal to the Word. In the former case, but not the latter, vocal cord spreading thus leads to a clearly audible gesture. It turns out that reduction of the consonant's glottal gesture is gradient, and as a result, even among the voiced consonants, some will be more voiced than others. So the assimilation is not categorical, in the sense that the glottal articulation for the target consonant is not necessarily the same as for the triggering voiced segments. It is simply that the listener's perception of the result is largely categorical.

3.1.3. The Role of Speech Rate in Such Cases

Both of these cases are gradient in that articulation depends on rate of speech. As rate of speech increases, articulations overlap more. In this voicing example, different articulatory targets on a single articulator come too close together, so one of them gets reduced in degree. In the first example, different articulatory targets on different articulators overlap, so one gets hidden even if it is not reduced. In both examples, an articulation can *sound* as if it has been entirely deleted.

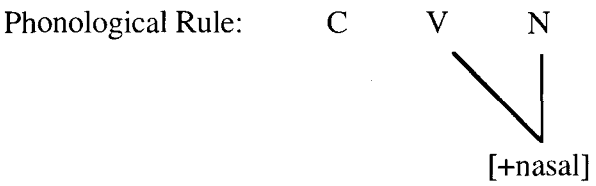
3.2. Assimilation of Unspecified Features

Next, we will consider two examples in which the segment targeted by assimilation is unspecified and the trigger context is specified.

3.2.1. Contextual Nasalization of Vowels

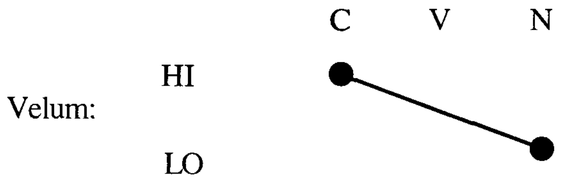
In many languages vowels are described as nasalized before nasal consonants. For example, the English word "bean" has a nasalized vowel [i]. A phonological rule of assimilation is shown in (12).

(12) Vowel nasalization as a phonological assimilation



Cohn (1990, 1993) investigated this in detail and showed that airflow from the nose in such words begins at vowel onset and builds up gradually over the vowel, indicating that the velum opens over the course of the vowel (as X-ray studies have also shown). The nasalization of these vowels looks very different from phonemic vowel nasalization in other languages, in which the vowel is nasal over its whole time interval. That kind of nasalization is seen in English between two nasals, as in "meany". On the other hand, "bean" looks very much like the mirror image of the nasalization of English vowels after nasal consonants, as in "need". If featural spreading is taken to occur only before nasal consonants, as in the typical rule, then the implication is that "bean" and "mean" are alike, but "need" is different. The phonetic implementation will then have to group these otherwise, because phonetically "bean" and "need" are alike and "mean" is different. The phonetic implementation makes more sense if instead we say that the vowel has no target of its own, and its degree and timecourse of nasalization are determined by interpolation from the consonants. This is schematized in (13): an oral consonant specifies a raised velum, the vowel does not care, and a nasal consonant specifies an open velum — so the vowel shows a C-to-N or N-to-C transition.

(13) Phonetic interpretation of vowel nasalization
Phonetic data: velum opening begins at vowel onset



A different view of the facts and their interpretation has been put forward by Bell-Berti, Boyce, and colleagues (Bell-Berti and Krakow 1991, Boyce et al. 1991). They present English examples in which non-nasal segments appear to have their own targets for velum position. Most notably, the vowel /ɑ/ has a lower velum than would be consistent with the schematic in (13) above; and /l/ appears to have a consistently lower velum than /s/ does. /s/ would be expected to have a very high velum position because of the phonetic constraints imposed by the feature value [+strident] (and to a lesser extent [-voice]). A strident fricative requires a high rate of airflow through the vocal tract, and therefore no leakage through the nose. Furthermore, for both /l/ and /ɑ/, the position of the tongue body doubtless influences the velum position. It is certainly well-known that the velum is lower when the tongue is lower, and such effects need to be factored in as part of phonetic implementation.

3.2.2. Vowel Allophony in Marshallese

Another, similar, example that I won't discuss in detail here is the allophonic fronting, backing, and rounding of the vowels of Marshallese by consonants with secondary articulations, described by Choi (1992). In brief, the short vowels of Marshallese contrast in height but not in backness or rounding. Variation in backness and rounding occurs, but as the predictable influence of consonants, which have contrastive lip and tongue body positions. (That is, Marshallese is unusual in using [Back] and [Round] as consonant, but not vowel, features.) Choi showed that the acoustic properties of the Marshallese short vowels are quantitatively predictable from the consonant context, and that no target values associated with backness or rounding need be assumed for them.

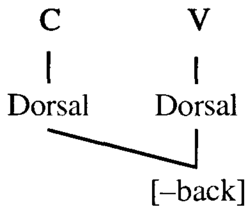
3.2.3. Velar Fronting

In many (if not most) languages, velar consonants are described as fronted before front vowels. For example, in English "key" the [k] is articulated on the hard, rather than the soft, palate. The position of the dorsal occlusion depends gradiently on the vowel. Assimilation to a front vowel can be formulated as a spread of the vowel's value for [Back]. Now if that were a complete account of the facts, this phonological rule would suffice; it would pose no problems in terms of phonetic implementation. The fact that the position of the tongue on the palate depends gradiently on that of the vowel would not stand in the way of a phonological assimilation rule: the rule says the consonant gets the same value as the vowel, and when the vowel is implemented in the front-back dimension, the consonant's implementation will simply follow from that. This is illustrated in (14).

(14) Velar fronting before front vowels as a phonological rule.

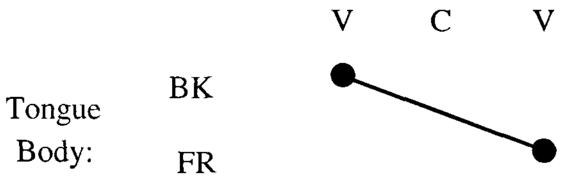
Example: "key" [k̟i]

Phonological Rule:



Such a phonological account of velar fronting is problematic for another reason: the fact that a preceding vowel's frontness also matters to the velar. As summarized in (15), in a vowel-velar-vowel sequence, the tongue moves from vowel to vowel during the velar consonant. Because the tongue forms both the vowels and the velar, the position of the velar articulation itself therefore moves during the consonant. The velar closure typically has no single location but depends entirely on the context. Spreading from the second vowel thus makes a wrong prediction about the realization of the velar. Instead, we say that the velar has no specification or target for backness (beyond the property of being dorsal), and let interpolation arrange its physical backness. (15) illustrates this. The back vowel and front vowel both have their tongue body positions specified, the consonant does not. (See Keating and Lahiri 1993 for further discussion.)

(15) Velar Fronting as phonetic interpolation



3.2.4. Phonetic Assimilation

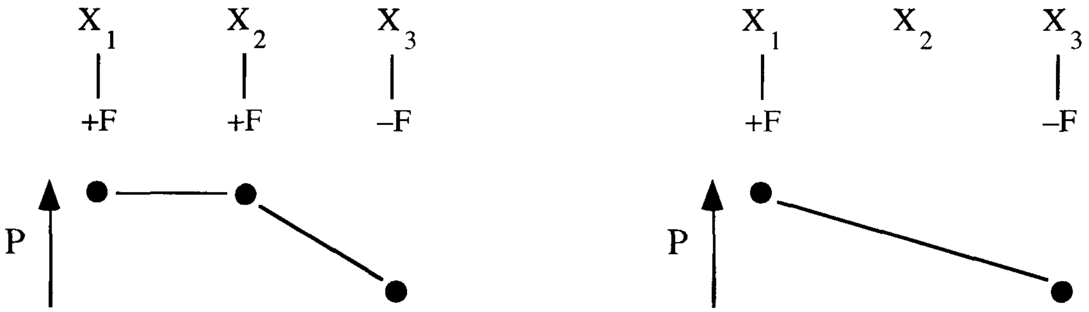
As already noted, the point of these examples is to show that "assimilation" need not be phonological (categorical); it can also be phonetic (gradient). It should be clear from these examples that gradient assimilation, due to overlap, reduction, or interpolation, is much more common than generally appreciated. This does not mean that sub-phonemic assimilation can never be categorical. Feature-spreading (or similar phonological operations of assimilation) in principle can produce categorical allophones, that is, derived feature combinations with targets different from those of the underlying segments. Nonetheless, this seems to be a rare occurrence. When suspected cases are examined carefully, they generally show gradient characteristics.

Consider the case of the many vowel allophones traditionally noted for Russian. Russian consonants have contrastive tongue-body articulations (palatalization/velarization) and these clearly have strong, gradient, effects on adjacent vowels. Most of the surface variation in Russian vowels is surely gradient, not featural. In Keating (1987) I showed that in general, variation in /i/ and /a/ can be accounted for in terms of gradient F2 target undershoot. For each vowel, though, there was one allophone that could not be attributed to target undershoot in this way. The vowel /i/ has a backed allophone target after non-palatalized consonants (this allophone is so well-established that it is considered a separate phoneme in many analyses of Russian); and the vowel /a/ has a fronted allophone target after palatalized consonants. That these are separate targets is seen pre-pausally, when the vowel has plenty of time to reach its usual target, but does not do so, instead reaching and holding the special target. It seems plausible that this is a case of phonetic feature-spreading: these special vowel targets could result from the spreading of [Back] from consonant to vowel. However, it is also possible that some version of the Window model (discussed below) could account for this variation.

3.3. The Role of Phonetic Underspecification

These examples thus illustrate what I call phonetic underspecification. This is a property incorporated into both Targets and Interpolation and Articulatory Phonology models. Phonetic underspecification means that not every segment has to have a specification, or target, for every feature/gesture. The particular consequences of this aspect of representation are different in the different models; I will limit myself here to Targets and Interpolation. Interpolation functions often do not care whether adjacent featural target specifications are from adjacent segments or not; they connect them up through an empty time interval between them, as illustrated in (16).

(16) Role of surface (phonetic) underspecification in such cases



The diagnostic for phonetic underspecification, then, is variability across contexts. If there is no phonetic specification, then what you see will depend entirely on the surrounding specifications, which will trigger interpolation through the unspecified span in a temporally-gradient fashion.

While contrast is an extremely important influence on variability, its influence can be quite indirect. We have already noted that a given feature can be realized along several parameters, so that several parameters can be somewhat influenced by a single feature value.

Furthermore, sometimes languages seem to specify aspects of sounds that are not determined by contrast alone. For example, English /l/ seems to be specified as having a tongue-backing ("darkening") component (Sproat and Fujimura 1993), though this cannot be attributed to contrast. For these various reasons, then, there is more phonetic specification than phonological specification. This point has been forcefully made (though in other descriptive terms) by Recasens (e.g. Recasens 1987). He notes that different consonant articulations place different phonetic demands on the tongue body; as a result, for some consonants the tongue seems hardly free to vary at all as a function of context, though it is not contrastively specified for tongue body features. Clearly, then, the theory of phonetic implementation needs to take all such effects into account.

3.4. Contrast and the Window Model of Phonetic (Under)Specification

Thus on this view, there can be a direct relation between phonological contrast and phonetic form. When there is contrast, there is feature specification, and therefore targets. The targets will be the main influence on the contour at that time. But when at some point in time there is no contrast that uses a given parameter, there will generally be no target at that time on that parameter, and the influence of context will obviously be strong. That is, contrast restricts contextual variability while lack of contrast gives rise to contextual variability.

I developed this idea as the "window" model of surface phonetics, summarized in (17) (Keating 1990). In this model, targets are not single points in space. Instead, they are spatial ranges of permitted values. You can think of these for articulation as constraints that say how much it matters how precise an articulation is. Some targets are very narrow ranges or windows; they permit little variation. Other targets are wide ranges or windows; they permit correspondingly more variation. In effect, windows turn phonetic underspecification from an all-or-none proposition to a gradient proposition. They allow an interaction between the constraints imposed by phonological contrast, and phonetic constraints such as mechanical linkages or aerodynamic requirements.

(17) Window model of phonetic (under)specification

Targets are ranges of permitted values, not single spatial points,
  – narrow window: a precise target; little variation across contexts
  – wide window: not a precise target; much variation across contexts
where window width is a continuous notion (narrower to wider)

In general, targets are narrow when in contrast, wide when not in contrast
Widest possible window = “phonetic underspecification” as in (15)

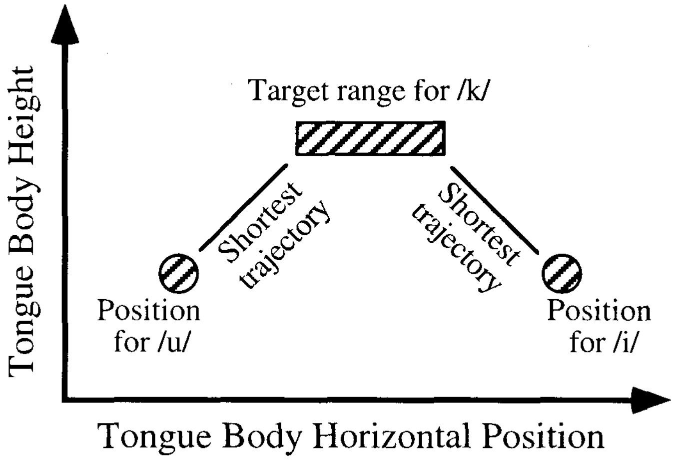
Interpolation through windows
  general idea = go as slow as possible, keep smooth trajectory

Looking at some acoustic data on Arabic uvular and pharyngeal fricatives in different vowel contexts has recently suggested to me that this proposal needs to be elaborated. The kind of window described so far is a range that must be reached no matter what — no matter how short the available time, no matter how great the velocity required, this is the absolute minimum that must be achieved. Yet clearly speakers often do more than this minimum when more time is available, as is the case for geminates and at slow speaking rates. It would seem that there is another, smaller, window within the larger window. This is the range that is preferred and which is reached when more time is available, but which is undershot when time is limited.

With target *ranges* rather than *points*, interpolation becomes a more complicated function. The general idea is to go as slow as you can while still making it into the required target range, but I had no precise proposals about this.

Guenther (1995) has implemented a neural net model of articulation that incorporates a windows-like idea. Articulatory targets are spatial ranges of acceptable positions, and these ranges function like my windows. Large target ranges give rise to shorter movements and more contextual variation. (18) schematizes one of his examples, which is velar position after /u/ vs. /i/, but which extends easily to velar position between /u/ and /i/. Guenther plots target ranges in two dimensions of tongue body position, Height vs. Backness. The velar stop /k/ has a limited range of possible heights but a wide range of possible backnesses, while the vowels in this illustration have limited heights and backnesses. The model makes the shortest connection between targets in this 2-dimensional space, so gets back velars after /u/ and front velars after /i/ and back-to-front movement between /u/ and /i/.

(18) Implementation by Guenther (1995) in a neural net model of articulation: tongue position for /k/ in /uk/ vs. /ik/



3.5. A Window-Style Model of Hyperarticulation

Guenther also develops his target ranges as an implementation of Lindblom's dimension ranging from hyper- (or *over*) articulation to hypo- (or *under*) articulation, under (19).

(19) Window-style model of hyperarticulation
shrinks targets
    slow speech (this is one attested strategy)
    careful speech
    prominence
    other prosodic conditions
reduces coarticulation (Lindblom 1990)

A small window is a kind of hyperarticulation because it requires more careful speech to reach the small target and it limits coarticulation. So slower speech, and more careful speech, would be modeled as a shrinking of target sizes, and faster/less careful speech as an expansion of targets. He also follows up on a result of deJong, Beckman, and Edwards (1993) that phrasal prominence results in a decrease in contextual variation and thus involves hyperarticulation. This can be modeled straightforwardly as some decrease in target range of the head of a prosodic domain. That is, depending on prosodic factors, we can shrink or expand a window. Thus depending on prosodic position, the degree of contrast to be observed in some case is smaller or larger. I believe that this approach (or something along these lines) will be very fruitful in accounting for a variety of prosodically conditioned variations: cases in which position means a narrower window (a strong specification) and other cases in which position means a very wide window (a weak specification, or near-neutralization).

4. Concluding discussion

Finally, I would like to place this discussion into the context of the distinction I have sometimes heard, that "phonology = formal, phonetics = functional". Such statements often are meant to imply that phonology and phonetics are two competing perspectives on the same phenomena, one of which must be better than the other. While there certainly are some individuals who believe this, the view outlined here is quite a different one. Phonology and phonetics are accounts of different phenomena — different levels in the grammar which carry out different computations. This is true in principle even when it is an unanswered empirical question whether any particular phenomenon belongs at one level or the other. Phonetics and phonology, formal and functional are orthogonal dimensions of inquiry and all combinations of them can be pursued.

Formal phonology: Probably most readers of this volume consider themselves to be in this category. Note that although almost all formal phonological theories incorporate at least some phonetic elements, such as phonetically-based distinctive features, they generally do not aim to account for (what are taken to be) phonetic facts.

Formal phonetics: This is a possible characterization (though not typical terminology) of the approach discussed here. The phonological part of the grammar produces symbolic representations which are passed to the phonetic component of the (formal) grammar. One consequence of the "Laboratory Phonology" movement has been to put this kind of work forward. The Conference in Laboratory Phonology is a conference which as of 1996 will be held in even-numbered years, alternating between the U.S. and abroad, and attended by phoneticians and phonologists; it is somewhat eclectic, but usually focused primarily on phonologically-informed phonetics. The proceedings are published (Kingston & Beckman 1990, Docherty & Ladd 1992, Keating 1994, Connell & Arvaniti in press). It has had a real impact on phonetics in that respect, perhaps a lesser one in phonology, but growing.

Functional phonology: By this is usually meant phonological theories or descriptions which are not formal in character, e.g. appealing to processing or learning constraints. These functional considerations may, but need not, be phonetic, i.e. pertaining to speech production or perception.

Functional phonetics: Presumably this includes much of traditional phonetics, focused on understanding the basic mechanisms of speech production and perception considered apart from linguistic structure. There is what might be called a "functional phonetics/functional phonology interface" which is concerned largely to explain typological generalizations about inventories.

However, there is no necessary restriction to one approach at a time, or to interface only formal-to-formal, functional-to-functional. For example, at UCLA phonologists are trying to base formal Optimality Theoretic constraints on functional phonetic considerations (e.g. Jun 1995, Hayes 1995, Silverman 1995, Kaun 1995, Flemming 1995, Kirchner 1995, Steriade in prep.). These studies highlight the importance for phonology of the properties of the speech medium.

Acknowledgment

My great thanks to Rob Hagiwara for help with ms. preparation, and to colleagues at UCLA and an anonymous reviewer for useful comments.

References

Bell-Berti, F. and R. A. Krakow (1991). Anticipatory velar lowering: a coproduction account. *J. Acoust. Soc. Am.* 90: 112-123.

Boyce, S. E., R. A. Krakow and F. Bell-Berti (1991). Phonological underspecification and speech motor organization, *Phonology 8*: 219-236.

Browman, C. and L. Goldstein (1989), Articulatory gestures as phonological units, *Phonology 6*: 201-251.

Browman, C. and L. Goldstein (1990), Tiers in articulatory phonology, with some implications for casual speech, in J. Kingston and M. Beckman (eds.) *Papers in Laboratory Phonology I: Between the grammar and physics of speech*, Cambridge University Press, pp. 341-376.

Browman, C and L. Goldstein (1992). Articulatory phonology: an overview. *Phonetica* 49:155-80.

Byrd, D. (1994), *Articulatory Timing in English Consonant Sequences*, UCLA Ph.D. dissertation distributed as *UCLA Working Papers in Phonetics 86*.

Choi, J.D. (1992), *Phonetic Underspecification and Target Interpolation: An Acoustic Study of Marshallese Vowel Allophony,* UCLA Ph.D. dissertation distributed as *UCLA Working Papers in Phonetics 82*.

Clements, G. N and S. Hertz (ms), An Integrated Representational System for phonology and acoustic phonetics, with a case study of English vocalic nuclei. Unpublished ms (1995).

Cohn, A. (1990), *Phonetic and Phonological Rules of Nasalization*, UCLA Ph.D. dissertation distributed as *UCLA Working Papers in Phonetics 76*.

Cohn, A. (1993). Nasalisation in English: phonology or phonetics, *Phonology 10*:43-82.

Connell, B. and A. Arvaniti, eds. (in press). *Papers in Laboratory Phonology IV* Cambridge University Press.

deJong, K., M. Beckman, and J. Edwards (1993), The interplay between prosodic structure and coarticulation, *Language and Speech 36*, 197-212.

Docherty, G. J. and D. R. Ladd, eds. (1992). *Papers in Laboratory Phonology II*, Cambridge University Press.

Flemming, E. (1995). *Perceptual Features in Phonology.* UCLA Ph.D. dissertation.

Fujimura, O. (1994). C/D Model: A Computational Model of Phonetic Implementation. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, Volume 17, pp. 1-20.

Goldsmith, John. 1976. *Autosegmental Phonology.* MIT Ph.D. dissertation.

Guenther, F. H. (1995), Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psych. Rev.*, in press.

Hayes, B. (1992), Comments on Nolan's paper, in G. Docherty and D. R. Ladd (eds), *Papers in Laboratory Phonology II,* Cambridge University Press, pp. 280-286.

Hayes, B. (1995). A phonetically-driven, Optimality-Theoretic account of post-nasal voicing. Talk presented at MIT, April 4, 1995.

Hertz, S. R. (1991). Streams, phones, and transitions: toward a new phonological and phonetic model of formant timing. *J. Phon.* 19(1): 91-109.

Jun, J. (1995), *A Gestural Approach to Assimilation in Consonant Clusters,* , UCLA Ph.D. dissertation.

Jun, S. (1993), Asymmetrical prosodic effects on the laryngeal gesture in Korean, Paper presented at Fourth Conference in Laboratory Phonology, to appear in proceedings edited by B. Connell and A. Arvaniti (in press).

Kaun, A. (1995). *The Typology of Rounding Harmony: an Optimality-Theoretic Approach.* UCLA Ph.D. dissertation.

Keating, P. A. (1987). Acoustic study of Russian vowel allophones. Poster presented at Spring meeting of the Acoust. Soc. Am. in Indianapolis. Abstract in *J. Acoust. Soc. Am.* 81 (Suppl.1): S67.

Keating, P. A. (1988a), The phonology-phonetics interface, in F.J. Newmeyer (ed.) *The Cambridge Survey, Vol.I: Linguistic Theory*, Cambridge University Press, pp. 281-302.

Keating, P. A. (1988b), Underspecification in phonetics, *Phonology 5*::275-297.

Keating, P. A . (1990), The window model of coarticulation: Articulatory evidence, in J. Kingston and M. Beckman (eds.) *Papers in Laboratory Phonology I*, Cambridge University Press, pp. 451-475.

Keating, P. A., ed. (1994). *Papers in Laboratory Phonology III.* Cambridge University Press.

Keating, P. A. and A. Lahiri (1993). Fronted velars, palatalized velars, and palatals. *Phonetica* 50:73-101.

Kingston, J. and M. E. Beckman, eds. (1990). *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech.* Cambridge University Press.

Kirchner, R. (1995). *Lenition in Optimality-Theoretic Phonology.* UCLA dissertation, to appear.

Krakow, R. A. (1989). *The Articulatory Organization of Syllables: a Kinematic Analysis of Labial and Velar Gestures*. Yale Ph.D. dissertation.

Lindblom, B. (1990), Explaining phonetic variation: A sketch of the H & H theory, in W. J. Hardcastle and Alain Marchal (eds.) *Speech Production and Speech Modelling*, Kluwer Academic Publishers, pp. 403-439.

Pierrehumbert, J.B. (1980), *The Phonology and Phonetics of English Intonation*, MIT Ph.D. dissertation.

Recasens, D. (1987). An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *J. Phonetics* 15: 299-312.

Silverman, D. (1995). *Phasing and Recoverability.* UCLA Ph.D. dissertation.

Steriade, D. (in prep.). Neutralization and the expression of contrast. UCLA ms.

Sproat, R. and O. Fujimura (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *J. Phonetics* 21: 291-311.

Zsiga (1993), *Features, Gestures, and the Temporal Aspects of Phonological Organization*, Yale Ph.D. dissertation.

Articulatory strengthening in prosodic domain-initial position

Cécile Fougeron [1, 2] & Patricia A. Keating [1]

[1] *Phonetics Lab, Dept. of Linguistics, UCLA*
[2] *Institut de Phonétique, Paris Sorbonne Nouvelle, URA 1027*

ABSTRACT

In this paper we show that consonants are articulatorily strengthened when they are initial in a prosodic domain. Linguopalatal contact for consonants in different prosodic positions was compared, using reiterant-speech versions of sentences with a variety of phrasings read by three speakers of American English. Four prosodic domains were considered, the Phonological Word, the Phonological (or Intermediate) Phrase, the Intonational Phrase, and the Utterance. Domain-initial consonants show more linguopalatal contact than domain-medial or domain-final consonants at the three phrasal levels but not the Word level. Furthermore, this strengthening is generally cumulative -- the higher the prosodic position, the more linguopalatal contact the consonant has. However, speakers differed in how many and which levels were distinguished in this way. In general, initial consonants in Word, Phonological Phrase, and a larger phrasal domain (Intonational Phrase or Utterance) were distinguished by linguopalatal contact. We suggest that this initial strengthening provides an alternative account for previously-observed articulatory declination of consonants. Turning to possible acoustic correlates of initial strengthening, we measured VOT, vowel duration, and stop burst energy. Some similarities between linguopalatal contact and these acoustic measures were found, but no strong correlations.

INTRODUCTION

The extent to which an articulatory constriction or opening is formed can be considered the magnitude or degree of that articulation. In the formation of an alveolar stop consonant, a smaller or larger closure can be formed depending on the magnitude of tongue raising and the orientation of the tongue. Several studies have shown that articulations of the tongue, lips, and velum are different in magnitude in word-initial position vs. non-initial position (e.g. Vaissière 1988, Krakow 1989, Browman and Goldstein 1992, Byrd 1994, Keating 1995). For example, Byrd (1994) showed that English alveolar and velar stops have more contact between the tongue and the palate when they are word-initial than when they are word-final. That is, word-initial position seems to enhance or strengthen articulations, increasing their magnitude. A few recent studies have shown that the glottal opening gesture associated with a word-initial voiceless consonant is greater in magnitude when the word is initial in a phrase. Pierrehumbert and Talkin (1992) showed that in English, an /h/ which is initial in an Intonational Phrase has more breathy noise relative to the vowel context (and presumably therefore a larger glottal opening) than does /h/ which is word-initial but phrase-medial. Similarly, they found that a phrase-initial /t/ has a larger VOT than a word-initial but phrase-medial /t/. Goldstein, in his published discussion of this paper (1992), cites similar results on aspirated stops from A. Cooper's (1991) transillumination data, concluding that increased magnitude of glottal opening is the cause of the VOT increase.

Subsequently, Jun (1993) compared the Voice Onset Time (VOT) of aspirated /p^h/ in Korean in three positions: when it was medial in a word (and therefore also medial in a phrase), initial in a word but still medial in a phrase, and initial in both word and phrase. The phrase that she used was the Korean Accentual Phrase, defined by the occurrence of a particular tonal sequence. Aspirated stops such as /pʰ/ in Korean typically have quite long VOTs, but Jun found that the VOTs of these stops varied depending on position. Figure 1 (a summary derived from Jun's Figure 6.2) presents her overall finding: /pʰ/ which is phrase- and word-initial had the longest VOT; /pʰ/ which is word-initial but phrase-medial had the next-longest VOT, and /pʰ/ which is word- and phrase-medial had the shortest (though still long) VOT. These results were consistent across the individual test utterances and speakers. Jun had no direct evidence that this VOT difference results only from the kind of increase in glottal magnitude demonstrated by Cooper (1991); an increase in the relative timing of oral and glottal articulations might also be involved. Nonetheless, whatever the cause(s) of Jun's results, they show that the position of a segment in both word and phrase influences some aspect of its glottal articulation.

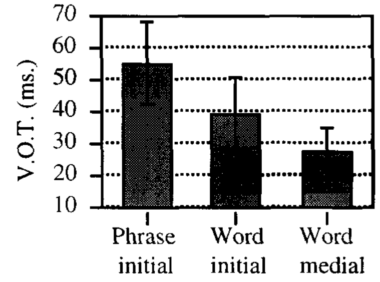


Figure 1. Summary results of Jun (1993), derived from Jun's Figure 6.2: VOT of Korean /p/ depending on position in word and phrase. The higher the position, the larger the VOT.

Our goal is to generalize and extend these previous results, in two ways. First, we want to establish whether a supralaryngeal articulator shows the same kind of phrasal-level strengthening effect as glottal articulation does. So far, only effects of position-in-word have been systematically examined for supralaryngeal articulations, as cited above. Since a range of supralaryngeal articulations are known to show strengthening in word-initial position, it is reasonable to expect that, like glottal opening, these articulations would vary above the word as well. Van Lieshout et al. (1995) compared lip emg activity for a rounded vowel in a word placed in either sentence-initial or sentence-final position, and found that the rounding activity was greater in the sentence-initial position. This finding is suggestive, but it might be due to something special about sentence-final, rather than sentence-initial, position. Here we consider a different supralaryngeal articulator, the tongue blade, in a variety of positions. We hypothesized that initial lingual articulations are strengthened compared to any non-initial lingual articulations.

Second, we want to establish whether more levels of phrasing can be distinguished by magnitude of articulation. Fujimura (1990) proposed that initial articulations in different prosodic domains are increased in magnitude. It is well-known that several levels of phrasing can be distinguished by segmental and tonal alternations, degree of final lengthening, pause, and perhaps other factors (e.g. Oller 1973, Beckman and Pierrehumbert 1986, Nespor and Vogel 1986, Ladd and Campbell 1991, Wightman et al. 1992, Jun 1993). These levels of phrasing, or prosodic domains, can be organized into a Prosodic Hierarchy (Selkirk 1984, Nespor and Vogel 1986, Hayes 1989, and Jun 1993 and Wightman et al. 1992 for comparative reviews). A Prosodic Hierarchy is a hierarchical organization of phrasal or prosodic levels into cumulatively larger domains. In this study, we test whether articulatory correlates mark edges of some prosodic domains. There is no reason to expect any one articulatory or acoustic correlate to distinguish all such prosodic domains; Wightman et al. 1992 found that pre-boundary vowel duration distinguished only four of the seven perceptually-distinct levels they tested. Therefore in this study we included four levels, to see whether any or all of them would be distinguished articulatorily. These levels, described further in the Methods section below, are the Phonological Word, the Intermediate or Phonological Phrase, the Full Intonational Phrase, and the Utterance.

Recall that the studies of glottal opening showed that its magnitude appears to be greater when a consonant is initial in a phrase (English Intonational Phrase or Korean Accentual Phrase) than when it is initial in only a word. An obvious question is whether this strengthening, like Prosodic structure, is hierarchical and cumulative. Does the magnitude of an articulation become progressively greater as its position moves up in the Prosodic Hierarchy? For example, will an articulation be more extreme when it is initial in an Utterance compared to a smaller phrase? We hypothesized that domain-initial strengthening occurs at every level of the Prosodic Hierarchy, and that it is cumulative: the higher the position in the hierarchy, the more strengthening an initial articulation in that position will show. Lingual articulations which are initial in higher domains should show more strengthening than those which are initial in lower, embedded domains.

A further motivation for our study was other recent work that at first glance might appear unrelated. Some researchers have proposed that articulations are more extreme earlier in utterances and decline gradually over the course of utterances (Vaissière 1986, Vayra and Fowler 1992, Erickson and Fujimura 1992, Krakow et al. 1994). So, for example, as we will describe in more detail later, Krakow et al. showed that the position of the velum during the voiceless stop /t/ (for which the velum is generally quite high) is higher early in a sentence and lower later in a sentence. They called this downtrend "supralaryngeal declination", on analogy with intonational, or fundamental frequency, declination (see Ladd (1984) or 't Hart et al. (1990) for reviews). Their finding that the velum is highest for a sentence-initial /t/ is consistent with our hypothesis of cumulative domain-initial strengthening, in that initial position in the sentence (prosodic Utterance) is expected to be the strongest position. However, it is not clear how to relate their finding of an *overall declination* to our hypothesized *domain-initial strengthening*. Experiment 1 below is a test of the initial strengthening hypothesis which allows us to distinguish between an overall decrease or declination vs. a local domain-initial increase in magnitude. We compare lingual articulations which are initial and non-initial in the different domains included in the study. With more than one non-initial articulation in a domain, we can test directly for declination. Both consonants and vowels are tested because declination has been found for both. Finally, Experiment 1 asks if initial strengthening affects the vowel in a CV syllable as well as the consonant, even though only the consonant is in absolute-initial position. In Experiment 2, preliminary in nature, we try to understand Krakow et al.'s declination in terms of possible phrase-level effects, on the hypothesis that particular phrasings could give rise to apparent declination. In this experiment we also look for some acoustic consequences of initial strengthening.

I. EXPERIMENT 1

A. Method

Linguopalatal contact as measured by Electropalatography (EPG) was used as the index of articulation strength. Substantial variation can be seen in the amount of linguopalatal contact for many consonants and vowels, and this variation indicates differences in overall oral constriction, making EPG an excellent choice for a study of variation in constriction magnitude. The Kay Elemetrics Palatometer was used; its pseudo-palates have 96 electrodes covering the entire hard palate and the inside surfaces of the molars. When the tongue touches an electrode, a circuit is completed and the Palatometer records the contact. Each sweep of the 96 electrodes takes 1.7 ms, and the sampling interval is 10 ms.

A1. Subjects and speech materials

There were three female American English speakers in the experiment, two naive about this study (Speakers 2 and 3) and one not (Speaker 1, the second author), each of whom already had a custom-made pseudo-palate for the Kay Palatometer. All are professional phoneticians in the UCLA Phonetics Lab. All normally produce alveolar rather than dental /n/.

To attempt to control for phrasing in our test stimuli, we designed sets of arithmetic statements in which the phrasing of the words would be crucial to conveying the meaning. The four different sentence types are presented in Table I. The first one is fairly flat, or free. The others require more careful disambiguation due to the occurrence of parentheses and the "times" operator early, middle, or late in the utterance. To control for the effects of lexical stress, different numerals (70, 100, and 89) were used in these sentences, each number having a different lexical stress pattern as shown in Table I.

<table>
<tr><th colspan="3">Model sentences:</th></tr>
<tr><td>1.</td><td>89 + 89 + 89 + 89</td><td>= a lot.</td></tr>
<tr><td>2.</td><td>(89 + 89) * (89 + 89)</td><td>= a lot.</td></tr>
<tr><td>3.</td><td>89 * (89 + 89 + 89)</td><td>= a lot.</td></tr>
<tr><td>4.</td><td>(89 + 89 + 89) * 89</td><td>= a lot.</td></tr>
<tr><th colspan="3">Numerals and stress patterns:</th></tr>
<tr><td>'σσσ</td><td>σ'σσ</td><td>σσ'σ</td></tr>
<tr><td>seventy</td><td>one hundred</td><td>eighty nine</td></tr>
</table>

TABLE I: Models for Experiment 1 corpus: 4 types of sentences, three numerals with different stress patterns.

However, so that all the syllables in the test sentences could be compared, regardless of the segments in them, we did not have subjects read these sentences in this form. Instead, we used reiterant speech, with the syllable "no" (/no/), for all of the utterance before "equals". For example, "(89+89+89) * 89 = a lot" was read as "(nonono no nonono no nonono) no nonono equals a lot". "No" was chosen because /n/ is an alveolar consonant whose linguopalatal contact is clear in EPG, and because /o/ is one of the American vowels which occurs in both stressed and stressless syllables, yet is quite different from /n/ in its contact pattern, thus making clear the consonant to vowel difference. Also, a sequence of /no/s is easy to say reiterantly. Larkey (1983) showed that while not all speakers can produce reiterant speech, those who can do so generally preserve prosodic properties of model utterances. All of our subjects, who were phoneticians, were able to produce good reiterant speech. Larkey found that not all speakers neutralize inherent durational correlates of the model segments, but this potential problem does not much affect our study, where duration was of only minor interest.

One subject (Speaker 1) read all combinations of the three numerals and four phrasings. However, because this is a large corpus it was preferable to record only a subset of it from the other subjects. Therefore two speakers (Speakers 2 and 3) produced all the numerals only in the flat phrasing (the first sentence type), and they produced the other phrasings only with "89". This number was chosen because its final primary stress should minimally affect the initial consonant. In fact, however, one speaker (Speaker 3) did put primary stress on the initial syllable of “eighty-nine”, with only secondary stress on the final syllable, in most of her tokens. We will return to this speaker difference below.

A2. Procedure and measurements

Subjects were not told or shown how to phrase each sentence, but instead were simply asked to convey the mathematical meanings indicated. All three subjects said they understood the sentences and could convey these meaning differences. Subjects practiced the sentences first with real words, then reiterantly a few times. Six good repetitions of each sentence were recorded, meaning that there was the right number of syllables and the prosody seemed to the speaker to convey the intended meaning. Few tokens were repeated on these grounds. The reiterant speech sounds about as natural as subjects' real-word models of these admittedly rather unnatural sentences. A constraint on file-length with the Kay Palatometer required subjects to produce these sentences one at a time at a moderate or fast pace. Several times a token had to be re-recorded because it had been cut off in the recording process.

Both the audio (12.8 kHz sampling rate, 16 bit resolution) and the EPG signal (100 Hz sampling rate) were recorded digitally in Kay Elemetrics's CSL format. EPG data were analyzed outside CSL by computing the percent of the 96 electrodes contacted in each data frame (Byrd et al. 1995). Each percentage point, then, is about equal to one electrode contacted. Figure 2 shows records for sample individual tokens of this "percent linguopalatal contact" over time, for the first two phrasings in the reiterant version of the number "89" produced by speaker 1. The distinction between the /n/s and the /o/s shows up clearly: the /n/s have more contact because they have greater oral constriction. Some /n/s have more contact than others, and contact for the /o/s is even more variable. Figure 2 also gives an EPG frame for two /n/s. The first one, labeled A, is taken from the initial syllable of the bottom sentence in the figure, indicated by point A. The second EPG frame, labeled B, is taken from the second syllable of this sentence, indicated by point B. Although there are only 45 electrodes active in the second /n/ vs. 59 for the first one, both /n/s still show a perfect alveolar seal for the stop closure. Much of the difference in contact is behind this seal, indicating a difference in tongue blade or body height between these two samples.

For each segment (stop or vowel) the extreme contact was measured from these records. Because the stop /n/ is characterized by a lingual-palatal constriction, its lingual articulation was measured as its percent linguopalatal contact in the frame with the maximum contact. Because the vowel /o/ is characterized by a lingual-palatal opening, its lingual articulation was measured as its percent linguopalatal contact in the frame with the minimum contact. With one data point measured per segment in the six repetitions of all the sentences, there was a total of 4320 measurements.

In addition, measurements of acoustic duration of /n/s and /o/s were taken from computer-displayed spectrograms.

Finally, a tonal transcription was made of each token by listening to the audio signal and looking at the f0 contour, the segmental durations, and any pauses. Pitch accents, phrase tones, and boundary tones were transcribed using the tonal part of the ToBI system (Silverman et al. 1992). These tonal transcriptions were used in determining the prosodic structure of each token, as described in the next section.

Statistical analyses were performed on the coded data using StatView (Abacus Concepts, 1992). The tests used are described in each results section below. We note here only that, because the speakers selected their own phrasings, the sample sizes of the different phrasal categories were rarely matched in these analyses (the number of tokens in a cell varies from 25 to 966); consequently the power of the different comparisons necessarily varies.

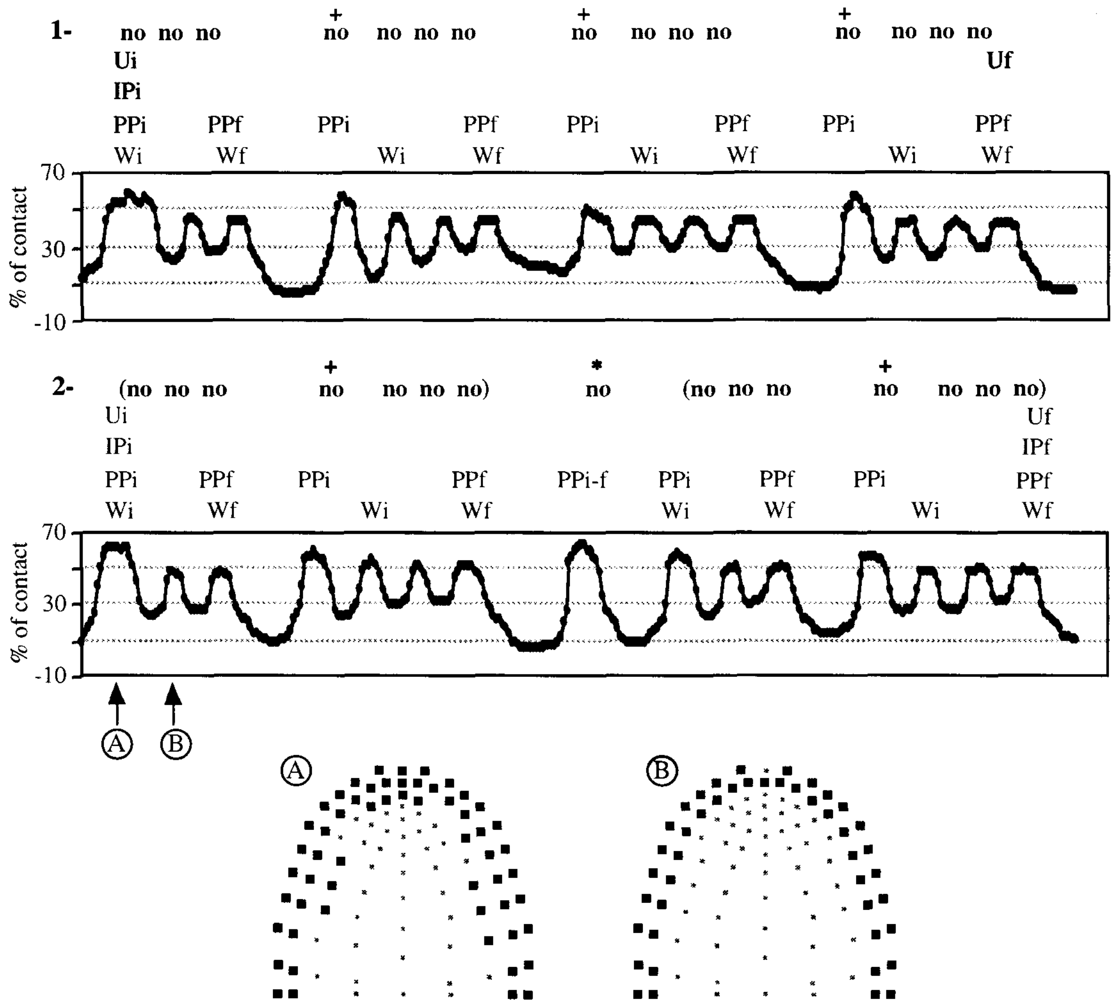


Figure 2. EPG sample data: Linguopalatal contact (as percent of electrodes contacted) over time for 2 samples of individual tokens of the sentences "89+89+89+89=a lot" and "(89+89)(89+89)=a lot" produced by Speaker 1. At bottom, electrode contact pattern at two timepoints indicated by "A" and "B" in second sentence.*

A3. Prosodic coding

Different theories of the Prosodic Hierarchy posit some kind of large phrase, some kind of phonological word, and two or more intermediate levels. In our study we considered four prosodic levels. Figure 3 summarizes the portion of a complete Prosodic Hierarchy that we considered. (1) Our lowest level is the Phonological Word (henceforth Word or W) which, in this study, will correspond to a lexical (morphosyntactic) word, e.g. a numeral like "eighty-nine". (2) Our highest level is an Utterance (henceforth Utterance or U). For some prosodic theories (Selkirk 1984) the largest prosodic domain is the Intonational Phrase, which is usually defined by the presence of a full intonational contour. For others (Nespor and Vogel 1986, Hayes 1989), two or more Intonational Phrases can be grouped into a larger domain that is called the Utterance. It is not our purpose here to argue a priori for or against the Utterance as a prosodic domain. However, as it seems possible that the very beginning of an utterance (in a pretheoretical sense, meaning simply when a speaker begins talking) could have some special status with respect to articulation, we do not want to preclude seeing its effects. Therefore we include the prosodic domain of the Utterance so as to distinguish the beginning of the Utterance from the rest. The data will show whether this distinction is justified. (3) This Utterance domain may contain one or more Intonational Phrases (henceforth IP, also called Full Intonational Phrases by Beckman and Pierrehumbert (1986)). An IP is defined by a complete intonational contour, including a final boundary tone. (4) Our intermediate level is a smaller phrasal domain, defined by at least one pitch accent and a phrase tone. It is the Intermediate Phrase of Beckman and Pierrehumbert (1986), and is roughly equivalent to the Phonological Phrase of other authors (Hayes and Lahiri 1991). For clarity of abbreviation we will refer to it here as the Phonological Phrase (henceforth PP).

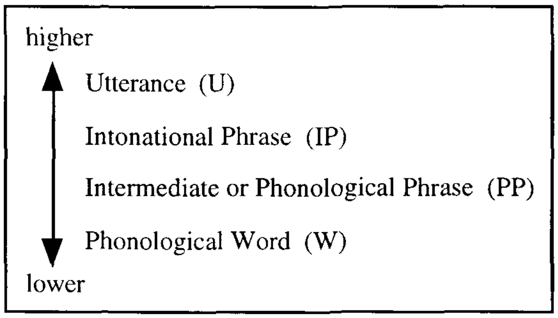


Figure 3. Portion of a Prosodic Hierarchy relevant to this study, ordered from the higher to the lower domains.

Two levels of prosodic structure were the same for all tokens in the corpus. First, the first syllable in each token was coded as U-initial, and all other syllables were coded as U-medial. There were no U-final test syllables, because the "equals a lot" portion of the tokens was not analyzed. Second, each numeral was a Phonological Word; the initial, medial, and final syllables of the numerals were coded as Word-initial, -medial, and -final, respectively. On the other hand, the operators ("plus" and "times") were not coded as Words -- of course they are Words, but not useful as such for our comparative purposes, because their single syllables are both W-initial and W-final. However, the operators were coded for all higher levels, and the very small number of operators which were both initial and final in a domain were counted as initial.

As was seen in Table I, the test sentences used parentheses to indicate the meanings of the expressions. We had expected these parentheses (and location of the "times" operator) to determine the intermediate levels of phrasing, such that we could assign the same prosodic structure to all tokens of each sentence type. This was not the case, however. There was variation in phrasing across, and often even within, speakers. Therefore prosodic structure was assigned token by token in accord with each token's tonal transcription. Recall that our two levels of phrasing are the Intonational Phrase IP, delimited by bigger breaks, a boundary tone and sometimes a pause, and the Phonological Phrase PP, delimited by smaller breaks and a phrase tone. The IP boundary tone indicates that there is f0 movement beyond what the phrase tone can account for. Therefore, a boundary tone was taken to indicate the end of an IP. The first syllable in each token was coded as IP-initial, and following syllables were coded as IP-medial or IP-final according to the transcription. Some tokens had no syllables coded as IP-final, meaning that the first IP continued past the last test syllable into the "equals a lot" part of the token. Similarly, a phrase tone was taken to indicate the end of a Phonological Phrase, and syllables were coded as PP-initial, PP-medial, or PP-final as appropriate. Both operators and numerals were coded for position in these two domains (as well as in the Utterance). If operators had been excluded from all analyses, there would not have been enough IP-initial or PP-initial tokens in the corpus. When an operator formed a domain by itself, it was included as initial in that domain.

Because prosodic domains are strictly hierarchical, the first syllable of each token was initial in every domain (Utterance, Intonational Phrase, Phonological Phrase, and Word); every IP-initial syllable was also initial in a PP, and if it was in a numeral, also in a Word; and every PP-initial syllable in a numeral was also initial in a Word. Similarly, IP-final syllables were also PP-final and, if in a numeral, also Word-final. In the statistical analyses, however, the codings for domain-initial syllables are used "exclusively" -- syllables are counted as initial only in their highest domain. For example, the IP-initial syllables are not also counted as PP-initial. In our analyses, "IP-initial" means "the highest domain in which this syllable is initial is the IP". Otherwise the analyses would be confounded by mixtures of levels.

Some of these codings, the domain-initial and -final ones, are illustrated in Figure 2. In general, the "times" operator generally formed a PP on its own, while the "plus" operator tended to form a PP together with the following numeral. There was a tendency for parentheses to correspond to IPs, though not in the example shown here.

So far we have discussed phrasing and coding in terms of syllables. The consonant and vowel of each syllable are coded the same as the syllable as a whole. Therefore what are coded as IP-final, PP-final, and Word-final /n/s are not in absolute final position in these domains; rather, the following /o/ is. Similarly, what are coded as Utterance-initial, IP-initial, PP-initial, and Word-initial /o/s are not in absolute initial position in these domains; rather, the preceding /n/ is. But we will refer to them according to their syllable, for convenience.

B. Results

B1. Domain-initial vs domain-medial and -final segments

The first analysis examined whether, within each prosodic domain, initial segments have greater contact than medial and final segments. For this analysis, all data from the three subjects were used. First, Utterance-initial segments were compared to U-medial segments, i.e. all other segments (recall that there are no U-final test segments). Second, IP-initial segments which were not also U-initial were compared to IP-medial and IP-final segments. Third, PP-initial segments which were not also IP-initial or U-initial were compared to PP-medial and PP-final segments. Finally, Word-initial segments which were not also PP-initial, IP-initial, or U-initial were compared to W-medial and W-final segments. This limitation to "exclusive" coding was necessary to ensure that initial segments in some smaller domain would not have greater average contact than the medials or finals because some of the initials were also initial in a larger domain.

Figure 4 shows the maximum contact for the /n/s and the minimum contact for the /o/s, in initial, medial and final positions for the four prosodic domains, for the three speakers separately. Table II gives the results of statistical comparisons by one-factor ANOVA and Fisher's PLSD post-hoc pairwise-comparisons, with .05 as the significance level for all tests.

Consider first the /n/s. Word-initial /n/ has significantly more contact than W-medial and W-final /n/ for Speaker 3, but not for Speakers 1 or 2. Speaker 2 does have significantly more contact for W-initial vs. W-medial, but W-initial is the same as W-final; Speaker 1 makes no distinctions within Words. PP-initial /n/ has significantly more contact than PP-medial and PP-final /n/ for all three speakers. IP-initial /n/ has significantly more contact than IP-medial and IP-final /n/ for all three speakers. U-initial /n/ has significantly more contact than U-medial /n/ for all three speakers. That is, except for the Word-level comparison, all domain-initial /n/s have more contact than their non-initial counterparts. Note that this result holds even though the prosodic coding was exclusive -- it is not the case, for example, that PP-initial /n/s as a group have more contact simply because some of them were also initial in some higher-domain. However, another possible confound in this kind of analysis could be that the higher domain-medial /n/s would have less contact simply because they would tend to also be medial or final in the Word. Nonetheless, when we compared initials, medials, and finals within the PP and IP levels, taking only /n/s which are Word-initial or in operators, much the same result was found ((b) in Table II). PP-initial /n/s still had significantly more contact than medials for two of the three speakers (Speakers 1 and 2); for Speaker 3 the direction of difference was maintained but it was no longer significant. IP-initial /n/s still had significantly more contact than medials for all three speakers. Thus there is apparent initial strengthening at the three phrasal levels tested, but not the Word level.

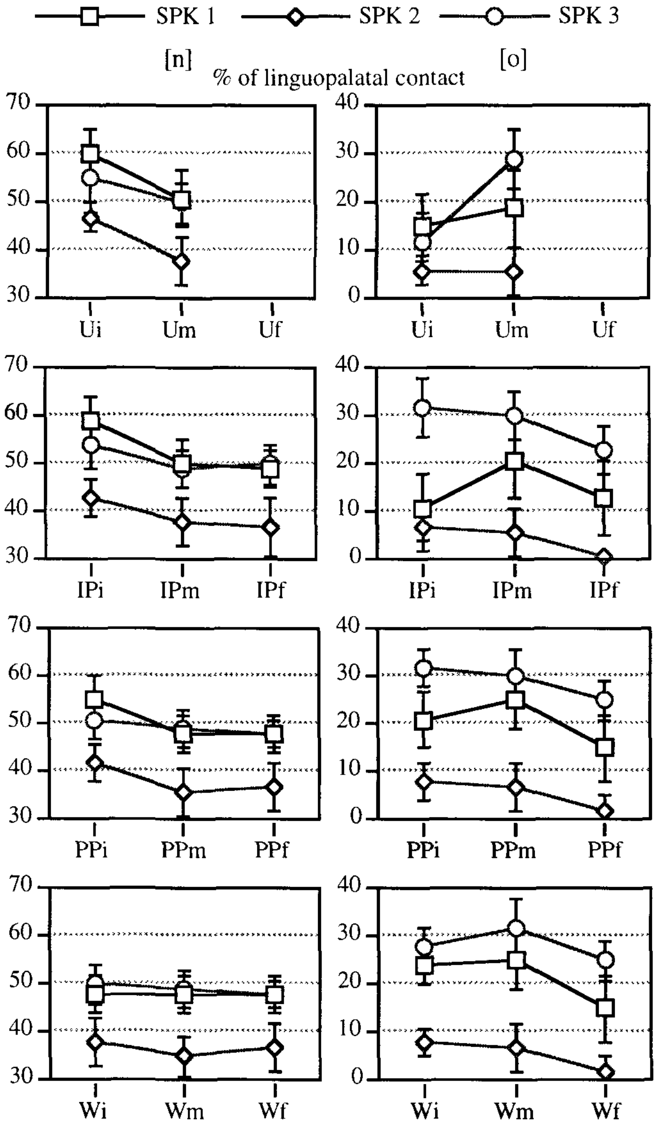


Figure 4. Maximum linguopalatal contact for /n/s (left) and minimum linguopalatal contact for /o/s (right) in three positions (initial, medial,final) in each of the four prosodic domains (Utterance, Intonational Phrase, Phonological Phrase, Word). Speaker results are shown separately within each panel. See Table II for significance of comparisons. Initial consonants generally show more contact, final vowels generally show less.

This domain-initial strengthening could be part of a larger declination within each domain, if medial /n/s also have more contact than finals. In Figure 4 the relation between these two positions can be seen to vary in size and direction across levels and speakers, and sometimes looks like a declination. However, the difference between medial and final might look like declination simply because more of the medials could be initial in some lower domain(s) and get strengthened at that lower level. This is likely to be the case because domain-final syllables are almost never initial in lower domains. Therefore domain-medial and domain-final /n/s which are not initial in any lower domain (Word or PP) were compared ((c) in Table II, except (a) for Word). There are three comparisons which give a significant difference between medial and final /n/s: at the Word and PP levels for Speaker 2, and at the IP level for Speaker 3. However, none of these differences indicate a declination, as in every case it is the final which has more contact than the medial.

The /o/s pattern rather differently. The most apparent pattern is that /o/s in domain-final positions show the least contact. We will interpret a decrease in contact for /o/ as meaning that it is more open, but tongue backing (de Jong 1995) could also contribute to the reduced contact. Under either interpretation, less contact means strengthening. Domain-final /o/s are more open than medial /o/s at every level for all speakers. And domain-final /o/s are more open than domain-initial /o/s, with Speaker 1's IP domain the only exception. No other differences pattern so clearly. Speaker 1 has a V-shaped pattern at every level: medial /o/s have more contact than either initial or final /o/s. Speaker 2 has similar contact for initial and medial /o/s except at the Word level, which shows a significant declining pattern. Speaker 3 has no pattern across domains: U-initial and W-initial /o/s have significantly less contact than medial ones, but PP-initial /o/s have significantly more contact than medial /o/s while IP-initial /o/s have the same contact as medial /o/s. Since PP-initial /o/s have more contact than medials, and since medial contact is greater than final in all domains, Speaker 3 has declining vowel contact within PPs. Across the speakers, then, for vowel articulations there is no consistent effect of initial position (bearing in mind that these vowels are never in absolute-initial position, only in the domain-initial syllable). Speakers 2 and 3 show some declination. However, as we would expect, vowels in final positions do show a consistent positional effect, with the least contact.

The only significant declining patterns in these vowel articulations are for Speaker 2 at the Word level and for Speaker 3 at the PP level. Thus there is no pattern in favor of overall declination in these results.

all numerals	[n]								[o]							
	U	IP			PP			W	U	IP			PP			W
spk1	(a)	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(a)	(a)	(b)	(c)	(a)	(b)	(c)	(a)
initial vs. medial	>*	>*	>*		>*	>*		ns	<*	<*	<*		<*	<*		<*
initial vs. final		>*			>*			ns		ns			>*			>*
medial vs. final		>*		ns	ns		ns	ns		>*		>*	>*		>*	>*
spk2																
initial vs. medial	>*	>*	>*		>*	>*		>*	ns	ns	ns		ns	ns		>*
initial vs. final		>*			>*			ns		>*			>*			>*
medial vs. final		ns		ns	ns		<*	<*		>*		>*	>*		>*	>*
spk3																
initial vs. medial	>*	>*	>*		>*	ns		>*	<*	ns	ns		>*	>*		<*
initial vs. final		>*			>*			>*		>*			>*			>*
medial vs. final		ns		<*	ns		ns	ns		>*		>*	>*		>*	>*

TABLE II: Results of statistical comparisons for Expt. 1 section B1. Ficher's PLSD post-hoc comparison of percent of linguopalatal contact between initial vs. medial vs. final [n] and [o]s at the four prosodic domain defined for the 3 speakers. (*) = $p < 0.05$; (ns) = $p > 0.05$. The column for each domain correspond to: (a) all data in Fig. 4, (b) = only W-initial and operators, (c) = not W-initial or PP-initial

Vayra and Fowler (1992) is one of the studies that has found supralaryngeal declination. In their study of Italian vowels, they report that only stressed vowels showed supralaryngeal declination, with a progressive centralization of stressed vowels from initial to final position in a trisyllabic CVCVCV word. This declination was found only when the words were produced in isolation, not Utterance-medially. In contrast, their unstressed vowels typically showed a V-shaped pattern, with medial vowels being more centralized (lower F1 for low vowels) than initial and final ones. This V-shape was not symmetrical, however; the initial unstressed vowels were more open than the final ones, consistent with an initial strengthening for unstressed vowels. Our results presented above were for stressed and unstressed syllables combined. Since Vayra and Fowler found declination only with their stressed vowels, it is reasonable to ask if the same result might be found with our stressed vowels. Figure 5 displays our data by syllable position, with the three speakers shown separately within each panel. In this figure, for Speakers 1 and 2 the W-initial (stressed) syllables are the initial syllables in "70"; the W-medial (stressed) syllables are the medial syllables in "100", and the W-final (stressed) syllables are the final syllables in "89". For Speaker 3, both the initial and final syllables of "89" were stressed, so both of these syllables were included here for this speaker. These comparisons, like Vayra and Fowler's, are across words. Word-initial here, as in the earlier analysis, is defined exclusively -- the Word is the highest domain in which these syllables are initial.

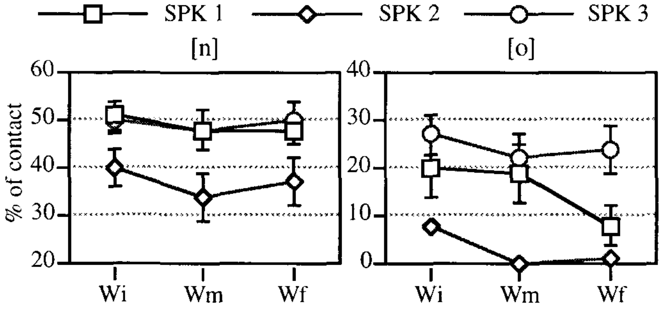


Figure 5. Stressed /n/ (left) and /o/ (right) contact in first, second, and third syllables of numerals. The three speakers are indicated as in the previous figure. See text for further description. There is no declination in either panel.

The results of this analysis of the stressed syllables are as follows. For /n/ (left panel), two speakers (Speakers 1 and 2) show initial strengthening. Speaker 3 has a V-shaped pattern: although W-initial /n/ has more contact than W-medial /n/, it is no different from W-final /n/. This is neither initial strengthening nor declination. For /o/ (right panel), again there are two patterns across the speakers. Two speakers (this time Speakers 2 and 3) show initial strengthening. Speaker 1 has final-lowering only. None of the three have any significant declination. In sum, our results are not like Vayra and Fowler's in that none of our subjects have declination of either stressed vowels or consonants. This could be because our numerals are in long utterances while their test words which showed this effect were in isolation.

In sum, the comparison between domain-initial and other /n/s within four prosodic levels provides evidence for initial strengthening in the three phrasal domains, but no evidence for declination.

B2. Hierarchical level of the domain boundary

Next we ask whether initial strengthening in different prosodic domains is cumulative. This analysis examined whether contact for initial consonants is greater when they are initial in higher prosodic domains than when they are initial only in lower prosodic domains. The comparisons were made by unpaired t-tests for every possible pairing of domain initial /n/s: W-initial /n/s which are not also initial in P, IP, or U; PP-initial /n/s which are not also initial in IP or U; IP-initial /n/s which are not also initial in U; and U-initial /n/s.

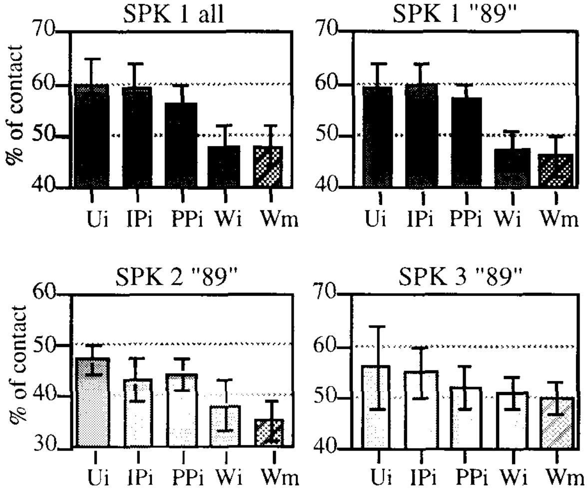


Figure 6. Maximum linguopalatal contact for /n/s which are initial in different prosodic domains (Utterance, Intonational Phrase, Phonological Phrase, Word, all shown as solid bars), vs. medial in Word (shown as striped bars). Speaker 1, all data, and all speakers, 4 sentences with numeral "89". See Table III for significant comparisons. In general, three levels are significantly distinguished.

Figure 6 shows the average contact for initial /n/s according to the consonants' highest domain, for all of Speaker 1's data and for just the sentences with "89" (the numeral for which all the speakers produced all the phrasings) for the individual speakers. To make an explicit comparison with the results of Jun's VOT study presented in Figure 1, we add here also the contact for word-medial /n/s, which are necessarily medial at every phrasal level. The figure shows that on average the amount of contact for /n/ is related to the hierarchical level of the domain boundary: higher levels show more contact, with Utterance-initial having the most contact and Word-medial the least. However, not all levels are reliably distinguished in this way, and although in almost every case three levels are distinguished, the speakers differ in which levels those are. Table III at the end of this section summarizes the significant differences.

In Speaker 1's data, in which all phrasings and all numerals are combined, three levels are distinguished: Utterance is not distinguished from IP, but all other domain-initial comparisons are significant. The result is the same when all three speakers' data are pooled; this congruence is not surprising, since Speaker 1's data dominate the pooled dataset.

When we consider the "89" data for each speaker, less consistent results are seen. Speaker 1's general pattern reported above does not quite hold for her "89" productions. Three levels are still distinguished, and W-initial, PP-initial, and IP-initial are still distinct, but the Utterance behaves slightly differently. U-initial has slightly *less* contact than IP-initial, but not significantly so, nor is it different from PP-initial. Speaker 2 distinguishes W-initial, PP-initial, and U-initial, but IP-initial is no different from PP-initial. Speaker 3 significantly distinguishes only two levels, with the lower levels W-initial and PP-initial distinct from the higher levels IP-initial and U-initial. IP-initial and U-initial are not different from each other. W-initial and PP-initial are also not different from each other, a result different from the other speakers. Speaker 3 is the subject who in the previous section was shown to have no significant difference between PP-initial and PP-medial when both are W-initial; this result is the other side of the same coin. Interestingly, the means of linguopalatal contact for Speaker 3 for all 4 levels in fact follow the trend of the other speakers and are consistent with cumulative strengthening, but fewer differences are significant.

In general, then, Word is distinguished from PP, and PP is distinguished from a higher domain, but whether that higher domain is the IP or the Utterance is variable. If the IP and Utterance domains were collapsed in these analyses, then the results would look much more consistent: Word, PP, and IP distinguished, with the exception of Speaker 3 for Word vs. PP.

Finally, the Word-medial data are included in Figure 6 for comparison. These W-medial /n/s are always syllable-initial. The within-domains analysis in the previous section already showed that Speaker 1 does not distinguish Word-initial from Word-medial /n/s. The other two speakers do, however. When this level of comparison is added to the phrasal levels just discussed, it turns out that two speakers miss one of the four distinctions that the third makes. Speaker 2 makes the most distinctions, distinguishing W-medial, W-initial, PP-initial, and IP/U-initial; Speaker 1 distinguishes the three phrasal levels but not W-medial vs. W-initial; Speaker 3 does not distinguish W-initial from PP-initial, but does distinguish W-medial from W-initial. Table III summarizes the significant results.

In sum, initial strengthening is cumulative over the prosodic domains considered. However, not all the domains are distinguished: Utterance-initial is not clearly a distinct level, and speakers vary in which levels they distinguish.

pooled data:	Ui = IPi > PPi > Wi > Wm
common subset:	Ui > PPi > Wi > Wm (IPi = PPi & IPi = Ui)
SPK1 all data:	Ui = IPi > PPi > Wi = Wm
SPK 1 "89":	IPi > PPi > Wi = Wm (Ui = PPi & Ui = IPi)
SPK 2 "89":	Ui > IPi = PPi > Wi > Wm
SPK 3 "89":	Ui = IPi > PPi = Wi > Wm

TABLE III: Hierarchical prosodic levels for /n/ significantly ($p<0.05$) distinguished by the amount of electropalatal contact. Results presented by subset of data analyzed.

B3. Relation between linguo-palatal contact and acoustic duration

The third analysis considered whether these differences in linguo-palatal contact could be ascribed to differences in duration. We have noted that final vowels are more open, and they are known to be acoustically longer. This relation could reflect less undershoot at longer durations (Lindblom 1963, Moon and Lindblom 1994). Oller (1973) has demonstrated acoustic lengthening of initial consonants. It is possible that consonant duration also determines contact undershoot: initial consonants might show more contact when and because they are longer. If that is the case, then our results on contact variation would be a side-effect of positional lengthening, certainly interesting in its own right, but not supporting our hypothesized strengthening. Therefore it is worth asking whether in our data consonants with more contact are also longer, and whether a correlation between contact and duration can be found. For this analysis, acoustic duration was compared with the maximum contact. For all speakers, the acoustic duration of all segments in the four phrasings for the 6 repetitions with "89", and in the flat phrasing for the 6 repetitions with "70" and "100", was used in a linear regression in which duration is the independent variable and the peak linguopalatal contact is the dependent variable.

When we examine the durations of the /n/s we do find the same pattern as for linguopalatal contact. Domain-initial /n/s are significantly longer than medial /n/s at Word, PP, and IP levels; Utterance-initial /n/s are shorter than U-medial /n/s for two speakers and longer for one. Domain-initial /n/s are longer than domain-final /n/s at all levels except for the Word level for Speakers 1 and 2, where they are equal.

However, despite this similarity of patterning within the two variables, the correlation between the amount of contact and the acoustic duration over this subset of the corpus was minimal (r^2=0.06). Nor does a correlation appear when we consider only domain initial (r^2=0.04) or final ones (r^2=0.006), or all the /n/s except initial ones (r^2=0.05) or except final ones (r^2=0.09). Therefore, the difference in the amount of contact of /n/s depending on their phrasal position is not related to their acoustic duration, no matter what the basis for comparison. It must be noted that degree of contact is not linearly related to opening; once the tongue loses contact with the palate, differences in opening are not reflected in contact. However, for these stop consonants there is always extensive contact, so this limitation of EPG cannot be blamed for the lack of correlation.

Looking at the vowels, on the other hand, we found a somewhat stronger correlation -- the longer the vowel is, the less contact (the more open it is) (r^2=0.3). So for vowels, duration is a contributing factor to the degree of opening, not a surprising result. Here the limitation of the EPG technique probably does matter; the correlation would likely be greater if open vowel positions were more reliably tracked by EPG.

B4. Declination of linguo-palatal contact?

The analyses in Section 1 above have already cast doubt on the occurrence of linguopalatal contact declination in the data, in that there is almost no reduction from medial to final /n/s at each level. Only two instances of declination were found for /o/, but no general declination emerged for either /n/ or /o/. This final analysis further considers whether, for the corpus as a whole, the amount of linguo-palatal contact declines across an utterance, either linearly over the whole utterance, or within smaller domains. This is done by testing for correlations between linguopalatal contact and linear position in the sentence: if there is declination, contact should decrease proceeding through the sentence. We tested for such declination by linearly regressing serial position of the segments (from syllable 1 to syllable 15) against their amount of contact, using all data from all speakers. We looked in various ways in which one might expect declination to be operating, but we did not find declination for any of them -- over all syllables (r^2=0.02), over stressed syllables (r^2=0.04), over Word-initial syllables (r^2=0.02), over PP-initial syllables (r^2=0.0006), and over IP-initial syllables (r^2=0.0003). Looking next for a within-domain declination, we compared only medial and final syllables, taking out the strengthened W-initial, PP-initial, IP-initial, and U-initial syllables. Again there is no declination (r^2=0.002). Similarly, no declination could be found for vowel contact, no matter which vowels were compared, for example either all vowels (r^2=0.01) or stressed vowels (r^2=0.004).

In conclusion, we can say that in our data there is no overall declination for either consonants or vowels.

B5. Summary of Experiment 1

The results of Experiment 1 can be summarized as follows. When initial, medial and final positions are compared within each prosodic domain, initial /n/s have greater contact than medial and final /n/s at the PP, IP, and Utterance levels; at the Word level this holds only for one speaker. When initial positions are compared across prosodic domains, three prosodic levels are distinguished for two speakers by significant differences in linguopalatal contact: Word, PP, and something larger. The other speaker distinguished only two levels: a higher level and a lower level. Finally, although isolated instances of declination did appear for vowels, there was no consonant declination for any speaker or at any prosodic level. We conclude instead that there is initial strengthening of lingual articulations at different prosodic levels.

II. EXPERIMENT 2

If initial strengthening is the only mechanism needed to account for our data, how then can we explain data showing apparent declination, such as in the Krakow et al. (1994) study? They looked at velum height during /t/ in /tn/ in reiterant versions of utterances of various lengths. The velum is always relatively high for /t/, but its exact height is variable. When the velum assumes a gradually more open position over a sentence, it is said to show a declination effect. When Krakow et al. compared the first and last words in their sentences, they found that the velum for the initial consonant of the first word was always higher than for the initial consonant of the last word. We also would expect this result, since the initial word is always initial in an Utterance and IP, while the last word is probably not. That is, we would expect to find this kind of difference in almost any comparison between the first word and a late word in a sentence.

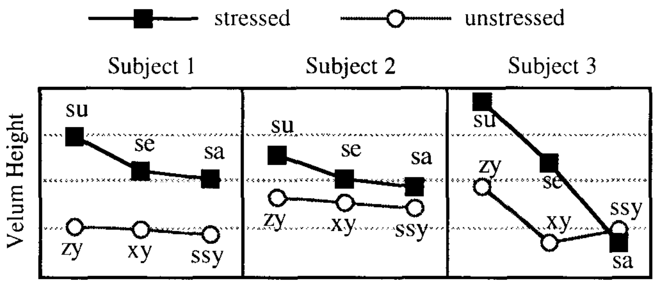


Figure 7. Results from Krakow et al. (1994) for velum height for /t/s in reiterant versions of underlined words in "Suzy saw sad sexy sassy Sid." and "Suzy saw sexy sassy Sid.", for their three subjects. Stressed initial syllables are indicated by filled squares; unstressed medial syllables by unfilled circles. Adapted from their Figure 5a.

In a further analysis, however, Krakow et al. found apparent declination over three words within the sentences. Figure 7, adapted from their figure 5a, shows velum height in /tEn/ for the reiterant versions of the three words "Suzy", "sexy", "sassy" in the two sentences "Suzy saw (sad) sexy sassy Sid". Comparing the initial (stressed) consonants of these three words, they found a main effect of word position for two of three speakers, and as can be seen in the figure, the general trend is similar for all three speakers. Although no post-hoc pairwise comparisons are reported, it does look as if velum height declines. Krakow et al. interpret this result as in Vayra and Fowler's (1992) study of vowels in isolated words: that articulatory position declines in stressed syllables from beginning to end of an utterance. Such an interpretation is seemingly at odds with our hypothesis of initial strengthening, since prosodic structure does not correspond to the linear order of syllables in a sentence. We offer instead a different interpretation of the Krakow et al. data: that the prosodic structure of their sentences was such that initial strengthening results in an apparent declination. If earlier syllables are initial in higher domains while later syllables are initial only in lower domains, then there will be progressively less contact over the course of a sentence. For example, if "sexy" is initial in a higher domain than "sassy", then "se" will be stronger than "sa". In Experiment 2 we consider on a small scale whether this interpretation of Krakow et al.'s results is plausible. This experiment is not intended as the definitive comparison of declination vs. initial strengthening, but simply as a demonstration that our account is a plausible alternative.

This experiment also looked for acoustic declination and for acoustic correlates of initial strengthening. As reviewed in the Introduction, VOT has been found to vary depending on phrasal position. Therefore, we used a voiceless stop test consonant and compared linguopalatal contact and VOT to see whether they vary with phrasal position in the same way in a single experiment. If they do, then at least two articulatory subsystems, tongue and glottis, would be shown to show prosodic strengthening in the same corpus.

The acoustic measures of interest were VOT and the level of energy at the release of the stop. We reasoned that if the amount of linguopalatal contact is larger for a given stop, then all else being equal, on average it will take longer to release that stop. Therefore the energy in the first milliseconds of the release of the stop will be smaller, and VOT will be larger. A larger VOT will also be found if the aperture of the glottis is larger in domain initial position, as seems likely (see review in Introduction). The acoustic measurements are also a first step towards determining if initial strengthening could have any perceptual salience.

A. Method

We could not repeat Krakow et al.'s study measuring velum position, but we did repeat it using EPG to measure linguopalatal contact during successive oral stops. The experiment was designed to manipulate the prosodic positions of these stops so as to produce or not produce apparent declination.

A1. Subjects and speech materials

One of our naive subjects (Speaker 3) from Experiment 1 served as the subject for Experiment 2. Like Experiment 1, Experiment 2 used reiterant speech, but with the reiterant syllable /tp/. Krakow et al.'s reiterant syllable /tn/ could not be used for this EPG experiment because the oral contacts for adjacent /n/ and /t/ would be indistinguishable in a reiterant sequence (/tntntn/).

Four test sentences, similar to the ones used in the Krakow and al. experiment, were constructed to vary the prosodic relationship between the test words "sexy" and "sassy": either they were likely to be at the same prosodic level or one was likely to be higher than the other. The test sentences are given in Table IV.

Model sentences:	
1.	Suzy saw sad sexy sassy Sid.
2.	Suzy saw sexy-sassy Sid.
3.	Suzy saw sexy, sad, sassy Sid.
4.	Suzy saw sexy, sassy-sad Sid.

TABLE IV: Models for Experiment 2 corpus.

The first sentence was Krakow and al.'s. In this sentence, "sexy" and "sassy" could combine into a single phrase, making "se" initial and "sa" medial in that domain. Other phrasings are also possible, and therefore in test sentence 2 we tried to encourage the phrasing of the two adjectives together by omitting "sad" and writing a hyphen between the two test words. In test sentences 3 and 4, we moved the non-test word "sad" around to try to force "sexy" and "sassy" into separate phrases, where each would be initial. Sentence 3 is expected to be phrased into two groups of two words each: (sexy, sad), (sassy Sid). In sentence 4, the hyphen between "sassy" and "sad" was expected to group them into one phrase, leaving "sexy" to form its own phrase: (sexy), (sassy-sad) Sid. The actual phrasings used by the speaker are presented below (Section B1).

A2. Procedures and measurements

The subject was not told how to produce any of the test sentences. She produced 6 repetitions of each of the four test sentences. This experiment was carried out during the same session as Experiment 1.

The percent of linguopalatal contact was calculated by the same procedure as in Experiment 1. It was measured at the maximum of contact for the /t/ in each syllable of the test words (reiterant versions of "Suzy", "sexy", "sassy"). In addition, four acoustic measurements were made using CSL. The VOT of the /t/ in /tp/ was measured from the waveform. The acoustic energy at the burst (in a 5 ms (64 point) window centered over the release of the stop) was measured from an FFT spectrum giving the RMS value over all frequencies. Finally, the duration of the vowel // was measured to provide an index of the local speech rate. VOT could be longer in some syllable simply because that syllable is longer.

A3. Prosodic coding

As in Experiment 1, four domains U, IP, P, and W were distinguished and syllables were coded for their positions in these domains.

B. Results

B1. Sentence prosody

The first question is what phrasing(s) the subject chose to use for the four sentences. Table V shows the sentences in Table IV but with the phrasing information. Parentheses are used to indicate boundaries of PPs while square brackets are used to indicate boundaries of IPs. It can be seen that the last sentence had no dominant phrasing for the six tokens. Based on these phrasings, then, sentences 1 and 2 are different from sentences 3 and 4 in that "se" has a higher prosodic position than "sa" in the first two sentences.

Sentences with phrasing:		
1.	[(Suzy saw)] [(sad) (sexy sassy Sid)]	(5)
2.	[(Suzy saw)] [(sexy-sassy Sid)]	(6)
3.	[(Suzy saw)] [(sexy, sad)], [(sassy Sid)]	(6)
4.	[(Suzy saw) (sexy), (sassy-sad) (Sid)]	(3)
	[(Suzy saw) (sexy)], [(sassy-sad Sid)]	(3)

TABLE V: Phrasing realized by subject in Experiment 2 for the 4 sentences in Table IV. In parentheses is indicated the number of repetitions (out of 6) using this phrasing. Intonational Phrases are indicated by square brackets, Phonological Phrases by parentheses.

B2. Linguopalatal contact depending on sentence type

Unlike in Experiment 1, the test sentences were first analyzed strictly by sentence type (1, 2, 3, 4 in Table IV) rather than by prosodic domain, since our interest is in a comparison with Krakow et al.'s results, where the actual phrasings are not known. There are not enough tokens in each condition to make statistical comparisons meaningful, so our discussion is limited to qualitative tendencies.

Figure 8 presents in the top row the amount of linguopalatal contact for the /t/s in the word-initial stressed syllables (filled squares) and the word-final unstressed syllables (circles) in the 4 sentences, averaging all repetitions of each, in the same format as the previous figure. (This figure also includes other information in the other rows that will be discussed below.) First, the most obvious result is that in all four panels the /t/ in the stressed, U-initial syllable "Su" always has the most contact. Second, the leftmost panel, for Sentence 1, shows the same pattern (for the tongue, not the velum) as Krakow and al.'s apparent declination. The /t/ for stressed “se” has a little more contact than the one for stressed “sa”. The contact for sentence 2, in the next panel, shows the same tendency. In this sentence "sexy-sassy" was combined into a single PP with "se" initial and "sa" medial. Therefore a contact difference is expected. Thus in these two sentences, apparent declination is seen. In contrast, the third and fourth sentences in the two panels to the right show about the same contact in "se" and "sa". There is no declining pattern across "Su", "se", and "sa" in these two sentences. It seems, then, that when the data are visually examined, declination can be made to come and go depending on the relative prosodic positions of the test syllables.

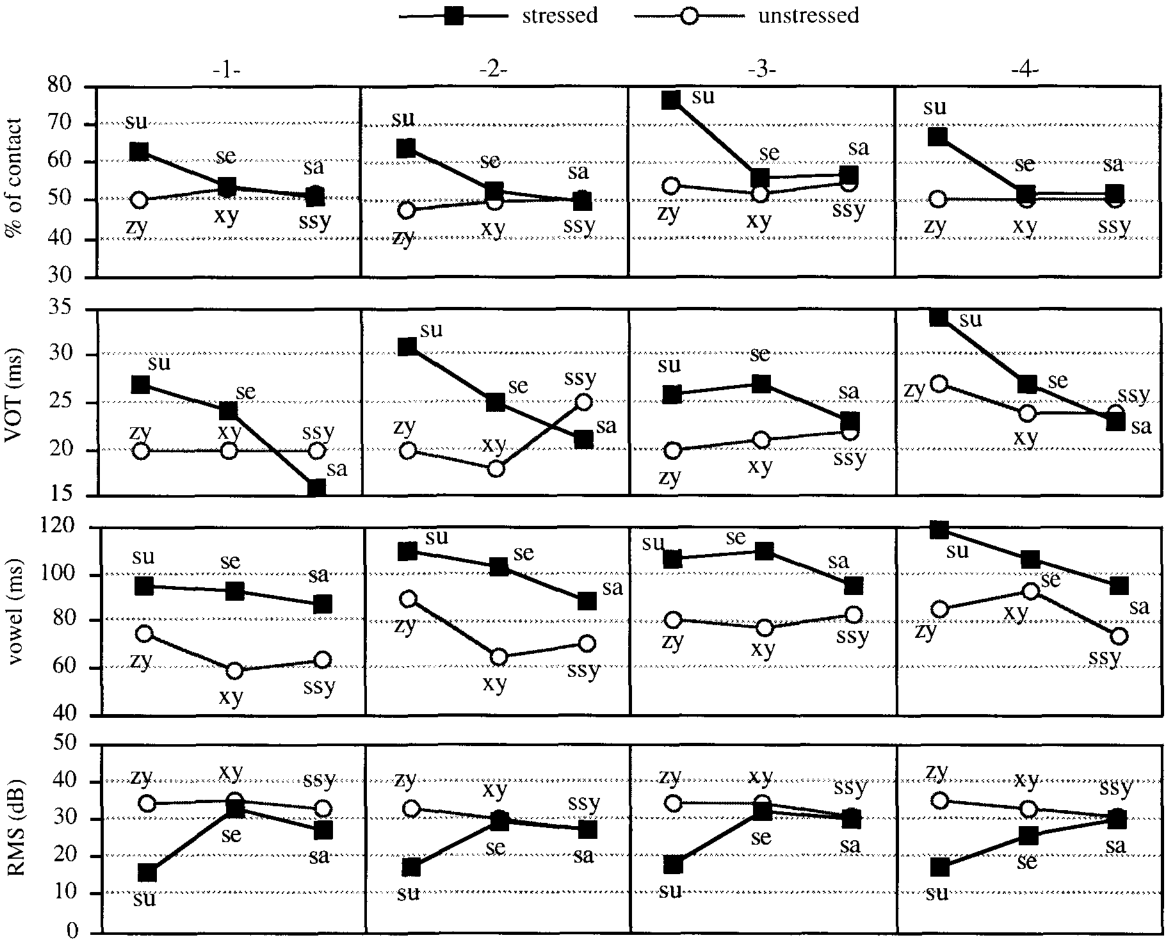


Figure 8. First row: Maximum linguopalatal contact for /t/s in reiterant versions of the same words as in Figure 7, each column corresponding to one of the four sentences listed in Table IV. Second row: VOTs of these. Third row: Duration of following //. Fourth row: RMS energy of /t/ bursts. Stressed initial syllables are indicated by filled squares; unstressed medial syllables by unfilled circles. Linguopalatal contact is seen to decline only in the first two sentences.

Next, consider the unstressed, non-initial syllables (circles) in the same three words. There is no difference between these /t/s as a function of their positions. We do not interpret this result as depending only on stress, as Krakow and al.'s or Vayra and Fowler's (1992) data might suggest. Instead, we suggest that it is equally important that these /t/s are not in initial position in the Word or any higher domain.

Finally, contact for both stressed and unstressed vowels was measured, though it is not shown in the Figure. It is basically the same in each of the three test words in the four sentences.

Summarizing these EPG results, the most apparent trend is that Utterance-initial stops had more contact than other stops. In addition, declination can be seen for sentences 1 and 2, where the prosodic strength of the test syllables decreases with linear order in the sentence, but not for the other sentences, where the syllables are prosodically comparable.

B3. Acoustic measures depending on sentence type

Figure 8 also shows, in the second through fourth rows, three kinds of acoustic data. Mean VOT after /t/, vowel duration, and mean energy at the /t/ release are shown for the Word-initial stressed syllables (filled squares) and for the Word-final unstressed syllables (circles).

a. VOT and vowel duration. All of the measured VOT values are on the low side for syllable-initial /t/, which is expected to be aspirated. This is presumably because the rate of speech in these reiterant sentences is fast. Within this range, however, there are clear differences among the three words in the sentences. In general, the VOT for initial /t/ in "Suzy" is longer than that for initial /t/ in "sexy", which is in turn longer than that in "sassy". The only exception is "Suzy" in sentence 3, where "Su" has a shorter VOT than "se", for which we have no explanation. Still. "se" has a longer VOT than "sa". Therefore regardless of the phrasing of "sexy" vs. "sassy" their VOT values do not seem to depend on their phrasal positions. VOT for stressed initial consonants, unlike linguopalatal contact, does seem to show declination between "se" and "sa" regardless of the phrasing. At the same time, the unstressed consonants in the same words show no trend of any kind. Thus the apparent sentence-level declination in VOT holds only for the stressed/initial consonants, similar to the results of Krakow et al. for velum height.

This finding of declination in VOT led us to check back in the earlier studies that had demonstrated phrasal effects on VOT (Pierrehumbert and Talkin 1992, Jun 1993), to make sure that their results could not be reinterpreted as declination (for example, if the higher-domain stops came earlier in their sentences than the lower-domain stops). This was not the case, however. In these studies, comparisons were made across sentences, that is, a W-initial consonant from one set of tokens was compared to an AP-initial or IP-initial consonant from another set of tokens, and the consonants were similarly located within the sentences. Therefore declination could not account for the differences; these differences must be due to the prosodic differences, as the authors suggested. In our study, consonants are compared within sentences, and declination is seen. Taken together, the previous studies and our study suggest that for VOT, both declination and initial strengthening are at play. This makes sense, because VOT depends not only on the glottal articulation, but also on the transglottal pressure immediately after release. Since we know that subglottal pressure does decline (e.g. Gelfer et al. 1987), it is entirely possible that subglottal pressure changes contribute in some way to VOT declination, whereas glottal opening shows initial strengthening effects, and that in our experiment the first effect was stronger.

Recall that our original question was whether VOT would vary with linguopalatal contact. We can see now that this is unlikely, since for stressed syllables in two sentences VOT declines while linguopalatal contact does not. Indeed, there is no correlation between the two measures (r^2=0.09). That is, the variation in VOT does not seem to be due at all to variation in constriction degree.

Finally, as seen in the bottom row of Figure 8, vowel duration measurements show a pattern very similar to that for VOT, again with sentence 3 different from the other sentences. Stressed vowels get shorter over the course of these sentences; it seems that overall speaking rate is increasing from beginning to end of a sentence. Therefore we asked whether the variation in VOT could be due to changes in speaking rate -- whether in stressed syllables VOT is correlated to vowel duration. There is however only a modest correlation (r^2=0.2). We conclude that variation in the local speaking rate is not the main reason that VOT declines within sentences.

b. Burst energy. The most consistent result with this measure was that the Utterance-initial /t/s have less energy than any other /t/s in the sentences. These /t/s also have more contact than any others. We had hypothesized that greater contact would give a longer release duration, hence less energy, and a negative correlation. Overall, the energy of /t/ is only modestly correlated with contact, (r^2=.3), with the correlation due mainly to the U-initial tokens having more contact and less energy than other /t/s. But while the U-initial energy is consistently low, the U-initial contact is quite variable, making it unlikely that contact is the main contributing factor in a stop's energy at release. The low subglottal pressure typically associated with Utterance-initial position could be an important independent factor, though we have no measure of this in this experiment.

Summarizing the acoustic analyses of this corpus, VOT and vowel duration declined over the three test syllables in three of the four sentence types. The RMS energy measure distinguished the Utterance-initial /t/ bursts as having less energy than other bursts. In these sentences, the U-initial /t/s also were distinguished by more linguopalatal contact, and linguopalatal contact gave the appearance of declination in two of the four sentence types. Thus there are similarities in the EPG and acoustic results, but closer analysis shows no strong correlations.

B4. Analyses based on prosodic position

In this section the data are analyzed as in Experiment 1 -- not by sentence type, but by initial prosodic position. The three stressed test syllables from all four sentences were grouped according to the highest domain in which they were initial, and the different domains were compared to see which, if any, are distinguished by linguopalatal contact and/or acoustic measures. These comparisons were made by unpaired t-tests. Figure 9 shows the maximum linguopalatal contact for the /t/s, the duration of the following vowels, the VOT of the /t/s, and the RMS energy of the /t/ bursts, in initial position in the four domains.

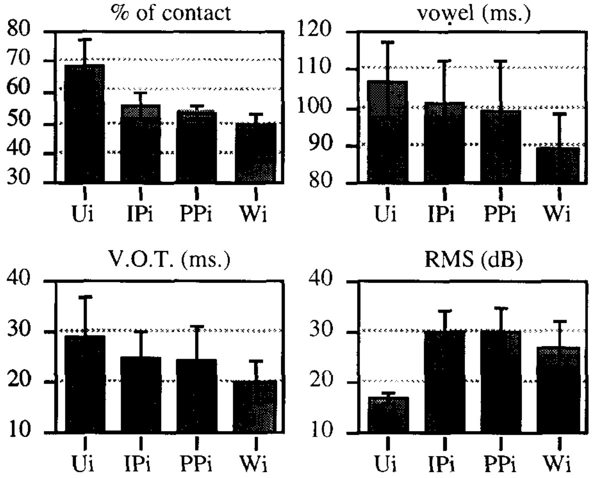


Figure 9. Maximum linguopalatal contact for /t/s (top left), duration of the following vowel (top right), VOT of /t/ (bottom left), and RMS energy of /t/ bursts (bottom right) initial in the four prosodic domains (Utterance, Intonational Phrase, Phonological Phrase, Word). In general, three levels are significantly distinguished.

Linguopalatal contact for /t/ distinguished three levels: U-initial, phrase-initial (IP and PP, not significantly different), and W-initial. The higher levels had more contact. Vowel duration distinguished the same three levels; the higher levels were longer. VOT for /t/ also distinguished three levels: U-initial, IP-initial, and W-initial; the higher levels had larger VOTs. Energy of the /t/ burst distinguished only U-initial from all others; U-initial energy was less. Thus U-initial syllables were reliably distinguished from other syllables by all measures, and the contact and duration measures all distinguished Utterance, Word, and some kind of phrase.

B5. Summary of Experiment 2

Experiment 2 was a preliminary one whose main purpose was to establish the plausibility of an initial-strengthening account of the supralaryngeal declination observed by Krakow et al. (1994). The results of this experiment support this account. The speaker used different phrasings for the different test sentences, and declination in linguopalatal contact is seen with two of the sentences, as expected. In these sentences the prosodic structure was such that initial strengthening would result in apparent declination; in the other sentences where the prosodic position of the test syllables does not decrease with linear order in the sentence, declination is not seen. However, further declination, not predicted by initial strengthening, is seen for two acoustic measures, VOT and vowel duration.

It does seem that prosody should be controlled or otherwise taken into account in declination experiments, as it is plausibly an independent factor. Further larger-scale experiments will be needed to determine whether initial strengthening alone can account for the apparent declination found in these studies.

We also found cumulative strengthening in Experiment 2. When linguopalatal contact is analyzed by prosodic position, differences are seen for U-initial vs. some kind of phrase-initial vs. Word-initial /t/s. This result is different from the result for this speaker in Experiment 1, where she made fewer distinctions than the other speakers, and different ones (W-initial and PP-initial < IP-initial and U-initial).

In this experiment, the relation between contact and acoustic measures is not strong. Therefore we can draw no conclusions about the acoustic consequences and perceptual salience of articulatory initial strengthening. We also therefore failed to show initial strengthening for linguopalatal contact and VOT within a single experiment.

III. GENERAL DISCUSSION AND CONCLUSION

A. Differentiation of prosodic domains

The present experiments allowed comparisons of four levels of prosodic structure: Word, Phonological (or Intermediate) Phrase, Intonational Phrase, and Utterance. Linguopalatal contact was compared both for different positions within each domain (to ask: Is there initial strengthening, rather than declination, within each domain?), and for initial positions across domains (to ask: Is this strengthening hierarchically cumulative?). The results for the within-domain comparisons can be summarized as follows. Domain-initial /n/s have greater contact than domain-medial and -final /n/s at the three phrasal levels, i.e. PP, IP, and Utterance; at the Word level this is so for only one of three speakers. Domain-final /o/s have less contact than other /o/s; there is no consistent effect of initial position on the vowels. That is, there is initial strengthening for consonants but vowels within prosodic domains. The results for the across-domain comparisons can be summarized as follows. Overall, three levels are distinguished by significant differences in linguopalatal contact: Word, PP, and something larger. But differences between speakers are noteworthy. Most embarrassing for the hypothesis, one speaker distinguished only two levels in Experiment 1; but the same speaker in Experiment 2 made a different, three-way, distinction. Thus while there is clearly cumulative initial strengthening across domains, it does not seem to be tied to distinguishing specific prosodic levels; we can say only that speakers distinguished two or three levels.

We mentioned in the Introduction that the domain Utterance is a controversial one, included in the present study just in case Utterance-initial position should emerge as special. The EPG comparisons of IP-initial vs. U-initial /n/s were quite varied across speakers, but in general these were not reliably distinguished by linguopalatal contact. In Experiment 2, the only subject did distinguish Utterance-initial from all other /t/s by all measures, so this domain cannot be dismissed entirely from consideration, but the evidence from this study for the prosodic status of the Utterance is weak.

Recall that Jun (1993) compared the Korean Word with a small phrase called the Accentual Phrase (AP), while Pierrehumbert and Talkin compared the English Word with the larger Intonational Phrase (IP), but no previous study looked at more than one kind of phrase at a time. Our study explicitly did this. Therefore it is worth asking how our levels of phrasing compare with these earlier studies. First, our Intonational Phrase is defined the same as Pierrehumbert and Talkin's, though unlike the IPs in their test sentences, ours are not so obviously syntactically based (and therefore more variable across speakers, most likely). Second, our PP, the domain of a phrase tone, is very similar to the Korean AP. It is generally thought that English does not have an Accentual Phrase (Beckman and Pierrehumbert 1986). Nonetheless, it is worth noting that almost all our PPs consisted of only one or two Words and had only a single pitch accent. Therefore our two levels of phrasing IP and PP are completely consistent with those in the previous studies. We found some initial articulatory strengthening at both levels in a single experiment. We also found similar differences in VOT as did Pierrehumbert and Talkin (1992) between IP and W, and as did Jun (1993) between PP and W.

Finally, we also cited studies that showed articulatory differences between word-initial and word-final consonants (where these lexical words were also prosodic Words). For example, Word-initial lingual consonants have been shown to have more linguopalatal contact and higher tongue blade positions than the same consonants in Word-final position. Typically, these experiments have compared consonants in CVC monosyllables, and from such comparisons it cannot be said whether the first consonant is strengthened, the last weakened, or both. The present experiment increased the number of consonants to be compared by putting three or more syllables within a domain. In this way we have shown that it is the initial consonant which is differentially affected. We speculate, following Krakow (1989) on the velum, that the same difference holds within syllables: that syllable-initial consonants are stronger than syllable-final ones. If this is so, then the word-level differences demonstrated by previous researchers in monosyllables confound Word and Syllable initial strengthening, making the effect quite large. The results of Experiment 1, in which two of the speakers had significantly more contact in W-initial vs. W-medial positions, suggest a difference between Word and Syllable levels, since our W-medial /n/s are syllable-initial.

B. Prosody and declination

Some of the analyses in this study were tests for declination in lingual articulation. For example, comparisons of domain-medial and domain-final /n/s tested whether domain-initial increases in linguopalatal contact could be ascribed to larger, declining, trends. These comparisons did find two cases of declination, but only for vowels: /o/ within PP for Speaker 3 and /o/ within Word for Speaker 2. These isolated cases were the only ones in Experiment 1, and no systematic declination emerged from several regression analyses of the pooled data. In Experiment 2, there was declination in contact only for the phrasings in which initial strengthening predicted it. At the same time, there was some declination of acoustic measures, VOT and vowel duration, for stressed syllables.

These experiments have shown that the position of a consonant or vowel in a prosodic domain can affect its degree of linguo-palatal contact, a measure which reflects its constriction degree. Most notably, a consonant which is initial in certain higher prosodic domains has more contact than does a consonant which is initial only in a lower domain, or is word-internal. Vowels which are final in a domain have less contact than non-final vowels in the same domain. Minimally, we have shown that in our corpus this prosodic conditioning is more important than any global, sentence-level downtrend or declination.

Beyond this, our results suggest a stronger claim: that apparent declination could be an artifact of the prosodic structure of typical short sentences, in which the beginnings of higher domains come before beginnings of lower ones. Articulatory declination is an appealing hypothesis, seemingly self-evident -- just as the clock gradually runs down, the speaker gradually runs out of energy. It is precisely because we took articulatory declination for granted that we undertook this study; we wanted to understand how it related to initial strengthenings. Nonetheless, despite its intuitive appeal, our results cast serious doubt on declination as an independent factor in speech motor control. It seems that the typical test sentences have been ones in which the predictions of declination and initial strengthening cannot be readily distinguished. On the other hand, initial strengthening cannot explain Vayra and Fowler's declination for stressed vowels in isolated trisyllables. An isolated Word forms an Utterance by itself so it could be that there is both U-initial strengthening and U-final weakening; in a three-syllable sequence the initial vowel would then be most open because of initial strengthening, and the final one least open because of U-final weakening. With the medial one intermediate in openness, the pattern would be a declination. This hypothesis cannot be tested with our corpus of trisyllabic numerals, because none of the words were in isolation.

It is interesting to mention in connection with declination a "failed" pilot study we conducted, looking for articulatory declination. We looked at long utterances of the form "my n̲et for X, my n̲et for Y,..." -- a repeated list would seem the perfect place to see declination. Yet the /n/s in "net" did not decline. Now we can appreciate that the crucial factor in the articulation of these /n/s is the flat prosodic structure of such lists; every "net" is exactly parallel to every other, so that we would not expect to see any differences in degree of strengthening.

Our results underline a point made by a few other researchers: the importance of understanding, controlling, and reporting the prosody of speech materials in articulation experiments. For the individual experimenter, unsought variation in prosody is a potential confound both within and across speakers, as our own experiment shows. And it can make comparisons across studies difficult or impossible, as researchers have always known. Yet, at the same time, awareness of prosodic differences between sentences can turn apparent random variation into predictable, lawful regularities of speech production, as we hope to have shown here.

This is not to say that prosody is easy to control, or that experiments on prosodic effects are easy to design. It can be especially difficult to find sequences of real words that can occur across a variety of prosodic boundaries and that contain segments appropriate to a given method of articulatory data collection (in our case, lingual consonants). Reiterant speech finesses this difficulty, but some subjects may not be able to produce it fluently (thought this was not a problem in the present study). Or doubts may arise as to whether reiterant speech induces some rhythmic properties that may enhance the prosodic phrasing, and thereby its manifestation in articulation. The present experiment does not address this point, and our results need to be confirmed with experiments using real words. Fougeron and Keating (1996) is an example of such a real-word design using normal sentences which, for one speaker of French, gives quite similar EPG results to our results here for English: a three-level distinction between Word, AP, and IP.

When future experiments are carried out to test for articulatory declination, it is important to bear in mind that declination could occur in many ways and at many levels. Not only could articulation decline across an Utterance, but also within each smaller domain. Indeed, f0 declination has been shown to be nested hierarchically within different-sized domains. There is declination within each smaller domain, with "resetting" at the beginning of each, and at the same time a larger-scale declination within each larger domain (Thorsen 1985, Maeda 1986, Ladd 1988).

If there is articulatory declination, it must be subject to such resetting: evidence for articulatory resetting at Word boundaries is found in Byrd (1994), in which the consonant at the end of a first word was weaker than the consonant at the beginning of the next word.

C. Demarcation of domain boundaries

The position of a segment within prosodic domains is its position relative to prosodic boundaries. Boundaries higher in the prosodic hierarchy are stronger boundaries. The results of our experiments indicate that segments in some prosodic positions are marked by articulatory strengthenings. We have focused so far on the idea that a given consonant will have more contact when it is initial in a higher domain than when it is initial in a lower domain, that is, a paradigmatic contrast. These results can also be considered in syntagmatic terms, that is, in terms of how prosodic positions affect sequences of segments within a single sentence. Consider a sequence of /n/s and /o/s; how different will each segment be from its neighbors? That depends on their relative phrasal positions.

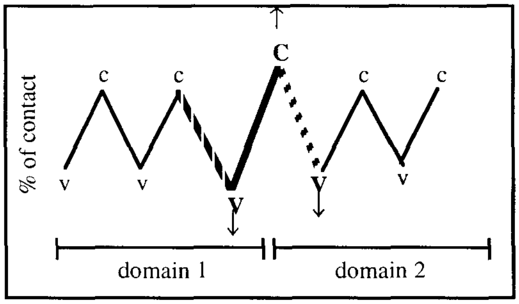


Figure 10. Schematic summarizing linguopalatal contact for segments spanning a boundary. Dashed line shows difference between last /n/ and last /o/ in a domain; bold line shows difference between last /o/ and first /n/ in next domain; dotted line shows difference between first /n/ and first /o/. Contrast in contact between last vowel and first consonant is greater than elsewhere in sequence.

Figure 10 is a schematic summarizing the data shown in Figure 4 in Section II. In a dashed line is the difference between the last /n/ and the last /o/ in a domain: this difference is large because the lengthened final vowel is quite open. In a bold line is the difference between that final vowel and the /n/ which is initial in the next domain: this difference is the largest one, because the final vowel is quite open and the strengthened initial /n/ has the most contact. In a dotted line is the difference between the initial /n/ and the following vowel: this also is large, due to the large contact of the /n/. This schematic shows that, at domain edges, the contrast between the consonant and the vowels in the sequence is enhanced in production. This articulatory enhancement of the contrasts within the sequence may contribute to marking the phrase boundary even more clearly than do the vowel or consonant alone. This need not have been the case; one can imagine that instead the more open final vowels would have pulled down the consonant, reducing its contact, so that the boundary-spanning segments would be overall more open, but with no greater difference between the consonant and vowel than in any other syllable. Instead, the vowel and consonant are more extreme in opposite directions.

What is the nature of this increase in contrast, whether viewed paradigmatically or syntagmatically? In stronger positions, lingual consonants increase their degree of contact or stricture. We speculate that initial strengthening makes stop consonants more stop-like (and therefore indirectly also more consonantal). The kinds of strengthening observed here and in studies of the glottis and velum could serve to reduce the acoustic energy of consonants, especially in the lower frequencies characteristic of voicing. These consonants would then stand out more in the acoustic signal, in which the overall level of energy is generally high. They would stand out not in the sense of being louder, but in being more different.

We have been non-committal as to whether vowels in initial CV syllables should be expected to show initial strengthening like consonants. On the one hand, perhaps only the initial segment is strengthened; on the other hand, perhaps the entire initial syllable is. In our data there is no domain-initial strengthening of vowels. We conclude that in a CV syllable only the C shows initial strengthening.

D. Conclusion

In conclusion, we have shown that amount of linguopalatal contact can depend in part on prosodic position. Overall, the three speakers in this study show increasing contact for consonants which are initial in the prosodic Word, small PP, and Intonational Phrase or Utterance. The acoustic or perceptual consequences, if any, of this articulatory variation remain to be established.

This variation in amount of contact means that the speaker marks initial boundaries of different levels of phrasing, and therefore *demarcates* these prosodic domains. Previous research on prosodic demarcation has focused on the ends of prosodic domains. Our results add to the much smaller literature on beginnings of domains. Not only glottal opening and velum closing, but also tongue blade articulations, are more extreme in domain-initial positions. We have shown that prosodic position affects even something that would seem to be an inherent property of a sound, namely its constriction. Thus higher-level prosodic organization is an important determinant of segmental articulation.

ACKNOWLEDGMENTS

This work was supported by an allocation MRT to the DEA de Phonétique de Paris to the first author and by NSF grant # SBR-9511118 to the second author. We wish to thank the subjects (Kimberly Thomas and Barbara Blankenship) for participating; Sun-Ah Jun and Sabrina Cooper for help with data analysis; Sun-Ah Jun for providing data from Jun (1993) and Janet Pierrehumbert for providing information about the sentences in Pierrehumbert and Talkin (1992); and Rena Krakow, Dani Byrd, Stefanie Shattuck-Hufnagel, and Jacqueline Vaissière for helpful comments and encouragement.

REFERENCES

Abacus Concepts (**1992**). *StatView* (Abacus Concepts, Inc., Berkeley CA).

Beckman, M., and Edwards, J. (**1994**). "Articulatory evidence for differentiating stress categories," in *Phonological structure and phonetic form: Papers in Laboratory Phonology III,* edited by P.A. Keating (Cambridge U. Press, Cambridge UK), Chap. 2, 7-33.

Beckman, M., Edwards, J., and Fletcher, J. (**1992**). "Prosodic structure and tempo in a sonority model of articulatory dynamics," in *Papers in Laboratory Phonology II: Gesture, Segment, Prosody,* edited by G. Docherty and D. R. Ladd (Cambridge U. Press, Cambridge UK), Chap. 3, 68-86.

Beckman, M., and Pierrehumbert, J. (**1986**). "Intonational structure in English and Japanese," *Phonology Yearbook* **3**, 255-310.

Browman, C. and Goldstein, L. (**1992**). "Articulatory phonology: an overview," *Phonetica* **49**, 155-180.

Byrd, D. M. (**1994**). *Articulatory timing in English consonant sequences.* UCLA dissertation, distributed as *UCLA Working Papers in Phonetics* **86**.

Byrd, D., Flemming, E., Mueller, C.A., and Tan, C.C. (**1995**). "Using regions and indices in EPG data reduction," *J Speech Hearing Res* **38**, 821-827.

Cooper, A. (**1991**) *Glottal gestures and aspiration in English.* PhD dissertation, Yale University.

de Jong, K. (**1995**). "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," *J. Acoust. Soc. Am.* **97**, 491-504.

Erickson, D. and Fujimura, O. (**1992**). "Acoustic and articulatory correlates of contrastive emphasis in repeated corrections," *ICSLP-92* **1**, 835-37.

Fougeron, C. and Keating, P. (**1996**). "The influence of prosodic position on velic and lingual articulation in French: Evidence form EPG and airflow data," to appear in proceedings of the ETR Workshop on Speech Production and Modeling, Autrans France May 1996.

Fujimura, O. (**1990**). "Methods and goals of speech production research," *Language and Speech* **33**(3), 195-258.

Gelfer, C. E., Harris, K. S., and Baer, T. (**1987**). "Controlled variables in sentence intonation," in *Laryngeal function in phonation and respiration*, edited by T. Baer, C. Sasaki, and K. S. Harris (Little and Brown, Boston MA), 442-35.

Goldstein, L. (**1992**). "Comments on chapters 3 and 4," in *Papers in Laboratory Phonology II: Gesture, Segment, Prosody,* edited by G. Docherty and D. R. Ladd (Cambridge U. Press, Cambridge UK), 120-124.

Hayes, B. (**1989**). "The prosodic hierarchy in Meter," in *Perspectives on meter,* edited by P. Kiparsky and G. Youmans (Academic Press, San Diego), 203-260.

Hayes, B. and Lahiri, A. (**1991**). "Bengali Intonational Phonology," in *Natural Language and Linguistic Theory* **9**, 47-96.

Jun, S. (**1993**). *The phonetics and phonology of Korean prosody.* PhD dissertation, Ohio State University.

Krakow, R. A. (**1989**). *The Articulatory Organization of Syllables: A Kinematic Analysis of Labial and Velic Gestures.* Ph.D. dissertation, Yale U.

Krakow, R. A., Bell-Berti, F., & Wang, Q.E. (**1994**). "Supralaryngeal declination: evidence from the velum," in *Producing speech: A Festschrift for Katherine Safford Harris*, edited by F. Bell-Berti and L. Raphael (AIP Press, Woodbury NY), Chap. 23, 333-353.

Ladd, D. R. (**1984**). "Declination: A review and some hypotheses," *Phonology (Yearbook)* **1**, 53-74.

Ladd, D. R. (**1988**). "Declination 'reset' and the hierarchical organization of utterances," *JASA* **84**, 530-544.

Ladd, D. R., and Campbell, W.N. (**1991**). "Theories of prosodic structure: evidence from syllable duration," *ICPhS XII* **2**, 290-93, Université de Provence.

Larkey, L. (**1983**). "Reiterant speech: An acoustic and perceptual validation,", *J. Acoust. Soc. Am.* **73**, 1337-45.

Lindblom, B. (**1963**). "Spectrographic study of vowel reduction," *J. Acoust. Soc. Am.* **35**, 1773-1781.

Maeda, S. (**1976**). *A characterization of American English Intonation*, Ph.D. dissertation, MIT.

Moon, S. J. and Lindblom, B. (**1994**). "Interaction between duration, context, and speaking style in English stressed vowels," *J Acoust Soc Am* **96**, 40-55.

Nespor, M. & Vogel, I. (**1986**). *Prosodic phonology* (Foris Publications, Dordrecht Netherlands).

Oller, D. K. (**1973**) "The effect of position in utterance on speech segment duration in English," *J. Acoust. Soc. Am.* **54**, 1235-47.

Pierrehumbert, J. and Talkin, D. (**1992**). "Lenition of /h/ and glottal stop," in *Papers in Laboratory Phonology II: Gesture, Segment, Prosody,* edited by G. Docherty and D. R. Ladd (Cambridge U. Press, Cambridge UK), Chap. 4, 90-117.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (**1992**). "TOBI: A Standard for Labeling English Prosody," *ICSLP-92* **2**, 867-870.

Selkirk, E. (**1984**). *Phonology and syntax: the relation between sound and structure* (MIT Press, Cambridge MA).

't Hart, J., Collier, R., and Cohen, A. (**1990**). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge Studies in Speech Science and Communication (Cambridge University Press, Cambridge UK), Chap. 5, 121-150.

Thorsen, N. (**1985**). "Intonation and text in Standard Danish," *JASA* **77**, 1205-1216.

Vaissière, J. (**1986**). "Comment on Abbs's Paper," in *Invariance and variability in speech processes*, edited by J.S. Perkell and D.H. Klatt (Lawrence Erlbaum Associates, Hillsdale NJ), 220-222.

Vaissière, J. (**1988**). "Prediction of velum movement from phonological specifications," *Phonetica* **45**, 122-139.

van Lieshout, P.H.H.M., Starkweather, C.W., Hulstijn, W., and Peters, H.F.M. (**1995**). "Effects of linguistic correlates of stuttering on emg activity in nonstuttering speakers," *J. Sp. Hearing Res.* **38**, 360-372.

Vayra, M. and Fowler, C. (**1992**). "Declination of supralaryngeal gestures in spoken Italian," *Phonetica* **49**, 48-60.

Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P.J. (**1992**). "Segmental durations in the vicinity of prosodic phrase boundaries," *J. Acoust. Soc. Am.* **92**, 1707-17.

VARIATIONS IN VELIC AND LINGUAL ARTICULATION DEPENDING ON PROSODIC POSITION: RESULTS FOR 2 FRENCH SPEAKERS.

Cécile Fougeron [1,2] and Patricia A. Keating [1]

[1] *Phonetics Lab., Dept. of Linguistics, UCLA*
[2] *Institut de Phonétique, Paris Sorbonne Nouvelle, URA 1027*

Abstract

Prosodic position, or the position of a segment in a prosodic constituent, is found to influence the articulation of the consonants [t] and [n] in French. Both lingual and nasal articulations for [n] can be affected by the position of the consonant, although the amount of linguopalatal contact seems to be affected more than the amount of nasal airflow. In general, the higher the Prosodic Position of the consonant, the less nasal flow [n] has and the more linguopalatal contact [n] and [t] have.

1. Introduction

Various aspects of prosody have been found to influence articulation. For example, the effect of lexical, phrasal or emphatic stress on articulatory gestures is relatively well documented [e.g. DeJong, 1995]. But another aspect of prosody is phrasing. By phrasing, we are referring to the grouping of segments into prosodic constituents and/or the marking of junctures between these constituents. Figure 1 gives an example of one of the possible phrasings of the sentence "La pauvre tata, Nadia et Paul n'arriveront que demain" (Poor Auntie, Nadia and Paul won't arrive until tomorrow). In this example, the sentence is structured into five prosodic levels that are listed on the left of the figure. These levels are organized hierarchically so that *Syllables* are grouped into *Words*[1], Words are grouped into *Accentual Phrases*, Accentual Phrases are grouped into *Intonational Phrases*, and finally Intonational Phrases are grouped together to form an *Utterance*.

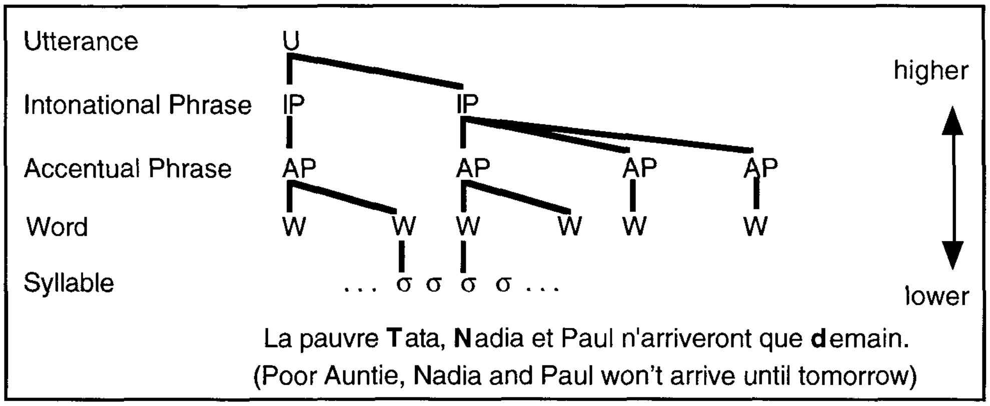


FIGURE 1: One of the possible phrasings of the sentence "La pauvre Tata, Nadia et Paul n'arriveront que demain."

This article is an extended version of the paper presented at the 1th ESCA Tutorial and Researsh Workshop on Speech Production and Modeling and 4th Speech Production Seminar, held in Autrans (France) May 20-24, 1996. A shorter version of this paper (including results for only one speaker) is published in the proceedings. We would like to thank here Richard Sico from Kay Elemetrics Corp. for his assistance and the ESCA commitee for the attribution of an ESCA grant to the first author to participate to the Workshop.

[1] "Words" here means only content words; function words will not be discussed since the test segments in the corpus are never preceded by function words.

What we call *Prosodic Position* is the position of a segment in this prosodic structure. For example, the [n] in "Nadia" is initial in a Word, in an Accentual Phrase, and in an Intonational Phrase. The [d] in "demain" is initial in a Word and in an Accentual Phrase, but is medial in the Intonational Phrase. The [t] in "Tata" is initial in a Word but is medial in the Accentual Phrase "Tata Nadia".

The best known effect of phrasing, or Prosodic Position, on articulation is final lengthening [Beckman, Edwards and Fletcher, 1992]. This lengthening occurs at the end of prosodic constituents like those shown in the tree of Figure 1. Our study examined the potential effects of Prosodic Position on articulation at the *beginning* of these prosodic groups.

The strength of the juncture between two prosodic constituents will depend on the prosodic level of the constituents: it will be stronger between two Accentual Phrases than between two Words within an Accentual Phrase. Therefore, a segment which is initial in an Accentual Phrase will be preceded by a stronger prosodic juncture compared to a segment which is initial only in a Word. Previous studies have shown that Prosodic Position influences the timing of articulatory movements and therefore the onset of coarticulation. In his cinefluographic study, MacClean [1973] showed that the strength of syntactic boundaries falling between the 2 vowels in a CVVN English sequence influences the timing of velic movement. Major syntactic (prosodically marked) boundaries, such as sentence or clause boundaries, delayed the onset of velic movement for [n] relative to the preceding vowel. On the contrary, for prosodically unmarked boundaries, the velic opening is initiated close to the begining of the articulatory movement of the first vowel. Similar results have been shown, although to a smaller extent, for tongue movement in a [kl] cluster [Hardcastle, 1985].

Prosodic Position influences not only the articulation of a segment relative to its neighboring context, but also influences the articulation of a segment compared to other segments in the speech chain (syntagmatic comparison) or in other sentences (paradigmatic comparison). For example, using the velotrace, Krakow [1989] examined the velic articulation of the consonant [m] in English as a function of its position in the Word and Syllable. She found that Word- or Syllable-initial consonants show a higher velum position than Word- or Syllable-final consonants. Several other studies have shown that articulations are weaker in Word- or Syllable-final positions compared to initial ones [among others, Benguerel, 1977; Fujimura, 1977; Vaissière, 1988; Byrd, 1994].

More recently, attention has been directed to the influence of position within higher prosodic domains, raising the question of whether particular changes in articulation may cue prosodic phrasing by marking prosodic boundaries. For example, in English [Pierrehumbert & Talkin, 1992] and in Korean [Jun, 1993], VOTs of aspirated stops have been found to be longer when the stop is initial in a large phrase (e.g. an Intonational Phrase), than if it is initial in a hierarchically lower prosodic domain. Most recently, Fougeron & Keating [1995, 1996] have shown that the strength of oral articulation depends on phrasal position. Measured with electropalatography (EPG), the amount of linguopalatal contact for [n] in English is larger in prosodic domain-initial positions (Word-, Minor Phrase- and Major Phrase-initial positions) compared to medial or final ones. This phenomenon is called "initial strengthening". Comparison of different prosodic levels shows that the degree of strengthening cues 2 or 3 levels of prosodic boundaries depending on the speaker.

The experiment presented here examines whether Prosodic Position influences the articulation of the consonants [t] and [n] in French. For the nasal consonant, prosodic effects are examined for both lingual and nasal articulations (in 2 separate experiments) in order to evaluate whether prosodic conditioning of articulation may be a general phenomenon in speech or is restricted to some articulatory subsystem. EPG offers us a direct measurement of the amount of linguopalatal contact, and airflow recordings give us an indirect observation of the velopharyngeal aperture.

2. Method

The amount of nasal flow and of linguopalatal contact was observed for the nasal coronal consonant [n]. In the airflow experiment, the consonant was placed in 2 vocalic contexts which vary in oral impedance: small for [a_a] and strong for [i_i]. In the EPG experiment, only the open vowel [a] was used in order to minimize the occurrence of linguopalatal contact that is not directly related to consonant articulation. The amount of linguopalatal contact was also observed for the oral coronal consonant [t] in the context [ɔ̃_ɔ̃].

The corpus consists of meaningful sentences where the test consonants were placed in 5 different Prosodic Positions. Table I gives the sentences designed for the sequence [ana] along with their prosodic labeling. Note that although prosodic boundaries are manipulated across the sentences, the test syllable remains in the same position in the sentences (5th syllable) in order to control for potential declination effects [Krakow et al., 1994]. In (1), [n] is initial in the highest prosodic group considered: the Utterance (Ui). It is separated from the preceding utterance by a long pause. In (2), [n] is initial in an Intonational Phrase (IPi). It follows the preceding Intonational Phrase that is defined on the right by a large pre-boundary lengthening, a long pause and a major pitch movement (here, a fall). In (3), [n] is initial in the lower prosodic domain defined by intonation in French, the Accentual Phrase [Jun & Fougeron, 1995] (APi). It follows the preceding Accentual Phrase marked by a small final lengthening without any pause and a minor continuation rise. In (4), [n] is medial in the Accentual Phrase "Tata Nadia" and is initial in the Word "Nadia" (Wi). In (5), [n] is Word medial and is initial in the Syllable (Si). Similar sentences were constructed for the test consonant [t] in the sequence "tonton" [tɔ̃tɔ̃] that allows quite productive word formation in French.

1- Utterance-initial (Ui):
"Paul aime Tat**a**. **Na**dia les protège en secret."
(Paul loves Tata. Nadia protects them in secret)

2- Intonational Phrase-initial (IPi):
"La pauvre Tat**a**, **Na**dia et Paul n'arriveront que demain."
(Poor Tata, Nadia and Paul will arrive only tomorrow)

3- Accentual Phrase-initial (APi):
'Tonton, Tat**a**, **Na**dia et Paul arriveront demain."
(Tonton, Tata, Nadia and Paul will arrive tomorrow)

4- Word-initial (Wi):
"Paul et Tat**a**-**Na**dia arriveront demain matin."
(Paul and Tata-Nadia will arrive tomorrow morning)

5- Syllable initial (Si):
"Tonton et A**n**abelle arriveront demain matin."
(Tonton and Anabelle will arrive tomorrow morning)

TABLE I: Stimuli designed for the sequence [ana]. Prosodic Positions are presented from highest to lowest in the examined prosodic hierarchy .

For the corpus with the consonant [n], nasal and oral flow were recorded directly into a Kay-CSL with a Rothenberg split mask. Variation in nasal airflow corresponds to variation in velum height as long as the total amount of airflow going to the nose is the same. We controlled for variation in the overall airflow between the test sentences by looking for variation in the acoustic energy of the following oral vowel, but all were comparable. Therefore, we will assume throughout this study that variation in nasal flow corresponds to variation in velic articulation. In a separate session, recordings of linguopalatal contact were made with the Kay-Palatometer for [n] in the context [ana], and for [t]. The custom made pseudopalate has 96 electrodes covering the entire hard palate and the inside surface of the molar. For these French-pseudo palates, the coverage of the electrodes extends toward the middle of the incisors which allows the dental articulation of French [n] and [t] to be captured.

Two Parisian French subjects, one female (1F, the first author) and one male (2M), were recorded for 10 repetitions of each sentence in the 4 corpora ([ana], [ini], [ana] EPG, [tɔ̃tɔ̃]).

Measurements were taken at the maximum of nasal flow in [n] and at the maximum of linguopalatal contact (in %) during the closure of [n] and [t]. Statistical analyses were made with unpaired t-tests (StatView) comparing the results by Prosodic Positions. A result obtained in one particular Prosodic Position is considered statistically different from the ones obtained in other positions if it shows a statistically significant difference from the other consecutive positions.

3. Results

3.1. Nasal flow experiment:

Figure 2 present the results for the female speaker (1F) combining the two vocalic contexts. Maxima of nasal flow in [n] are plotted by Prosodic Positions from the highest to the lowest prosodic level. For this speaker, the amount of nasal flow in [n] is affected by the Prosodic Position of the consonant. Three out of five Prosodic Positions are significantly distinguished by the amount of nasal flow and the tendency is for nasals in higher Prosodic Position to be marked by less nasal flow than nasals in lower Prosodic Positions. Thus, [n]s that are initial in a Word have more nasal flow than [n]s that are initial in an Accentual Phrase, which in turn have more nasal flow than the [n]s initial in an Intonational Phrase. The Utterance initial and Syllable intial positions do not follow this trend and are not significantly distinguished by nasal flow. These different patterns are indicated by the shading on the figure.

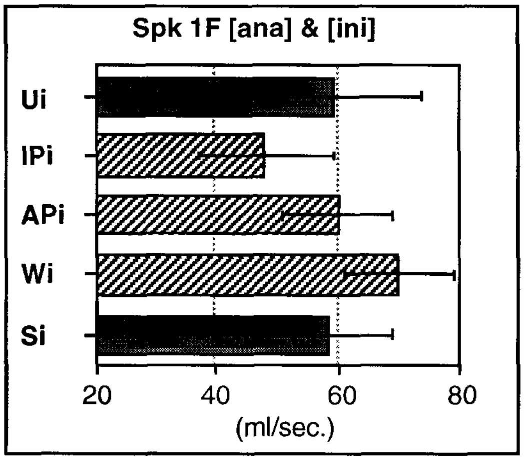


FIGURE 2 : Maxima of nasal airflow in [n] depending on its Prosodic Position. Data for speaker 1F grouping the two vocalic contexts [ana] and [ini], averaged over 10 repetitions each.

For the male speaker, the results are not so clear. When the two vocalic contexts are combined, none of the Prosodic Positions are distinguished by the amount of nasal flow. Figure 3 present the 2 vocalic contexts separately. For the sequence [ana], only one position stands out significantly (Accentual Phrase-initial position), but the others are not different. For the sequence [ini], there is a two way distinction between the phrasal levels (APi, IPi, Ui) and the word levels (Wi and Si). Again, the higher levels have less nasal flow than the lower ones.

In sum, the influence of Prosodic Position on the amount of nasal flow tends to vary depending on the speaker; both in how many Prosodic Positions are distinguished if any, and what these positions are. But when the amount of nasal flow is systematically affected by Prosodic Position, the higher the position, the less the nasal flow.

Similar results have been found in another experiment where the amount of nasal flow was observed for 4 French speakers in a slightly different corpus [Fougeron, 1996]. In that study, different Prosodic Positions were distinguished by a significant variation of nasal flow for 3 out of the 4 speakers. The speakers varied in the nature and number of Prosodic Positions that were distinguished but the tendency was the same as that observed in the present study: nasals in higher Prosodic Position are characterized by less nasal flow.

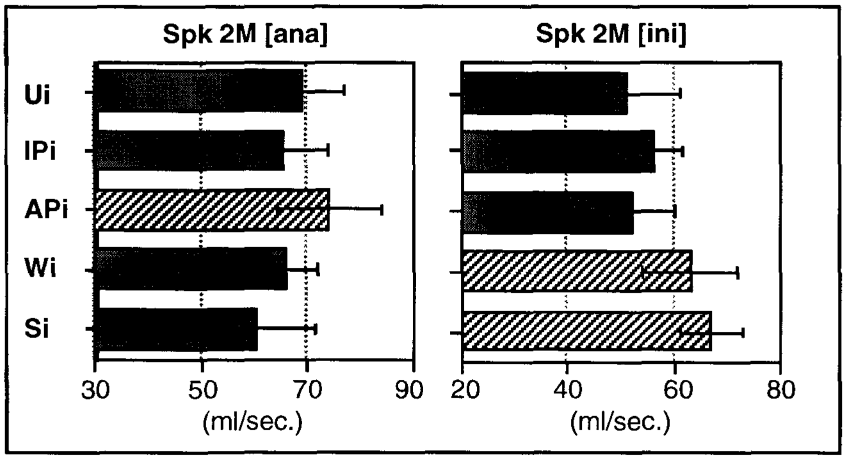


FIGURE 3 : Maxima of nasal airflow in [n] depending on its Prosodic Position and separated by vocalic context. Data for speaker 2M, averaged over 10 repetitions.

3.2. EPG experiment:

The maximum percent of linguopalatal contact for the consonants [n] and [t] is plotted against Prosodic Position in Figure 4 for speaker 1F (top) and speaker 2M (bottom). For the consonant [n], 4 out of the 5 Prosodic Positions are distinguished by the amount of linguopalatal for both speakers. The [n]s in higher Prosodic Positions have more linguopalatal contact than the [n]s in lower Prosodic Positions. The only positions that are not distinguished by linguopalatal contact are the Intonational Phrase-initial and Utterance-initial positions. For the consonant [t], the speakers make the same distinctions as for the consonant [n] but the difference between the Word-initial and Syllable-initial positions is not significant for speaker 2M.

Figure 5 gives an example of the pattern of linguopalatal contact for [n] at different Prosodic Positions for one repetition of speaker 1F. All consonants have a full closure and most of the variation in the amount of contact is behind the seal. This increase in linguopalatal contact as [n] occupies a progressively higher position in the prosodic hierarchy may be due to a modification in tongue blade orientation and/or tongue body height.

In sum, the amount of linguopalatal contact varies depending on the Prosodic Position of the consonants: the higher the position, the greater the amount of contact. This variation in linguopalatal contact allows the distinction of several Prosodic Positions: at least 3 out of the 5 positions for the 2 speakers and the 2 consonants observed. In our previous experiment on English [Fougeron and Keating, 1996] using a quite different style of corpus, we found fewer Prosodic Positions distinguished by the amount of linguopalatal contact. The 3 speakers varied both in how many positions were distinguished and what these positions were. However, in common with the present experiment, we generally found that Intonational Phrase- and Utterance-initial positions have the same amount of linguopalatal contact, whereas the distinction between Accentual Phrase- and Word-initial positions is fairly consistent.

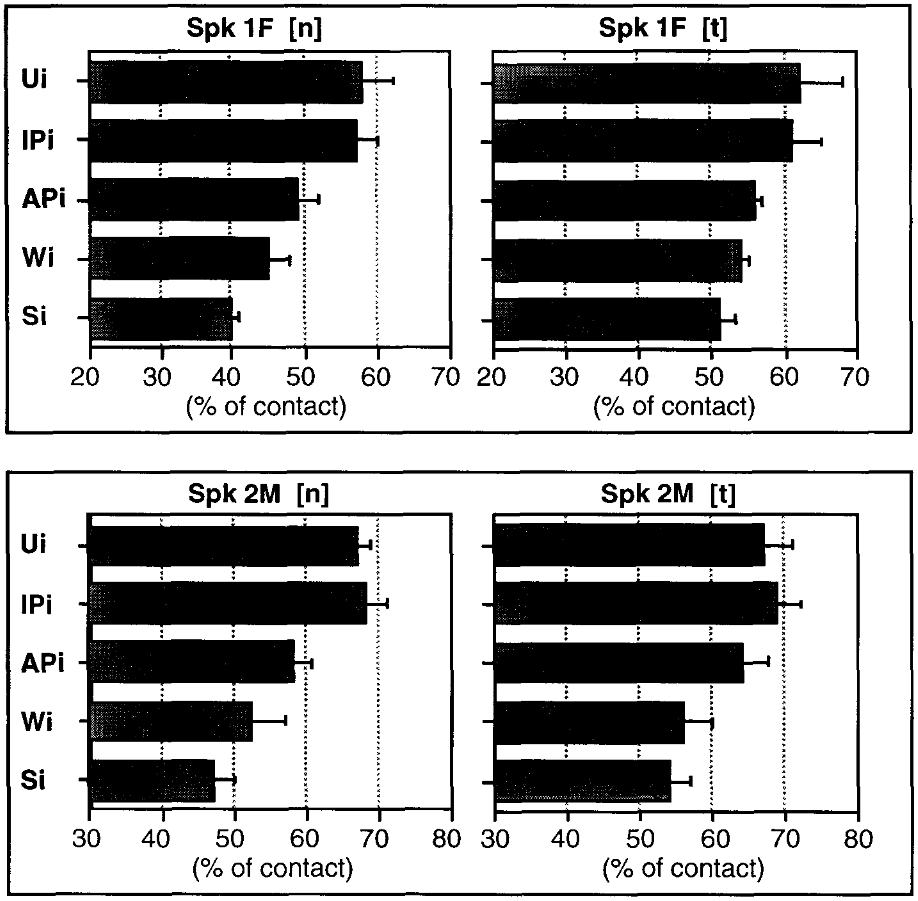


FIGURE 4: Maxima of linguopalatal contact in [n] and [t] depending on their Prosodic Positions. Data for speaker 1F (top) and speaker 2M (bottom), average over 10 repetitions.

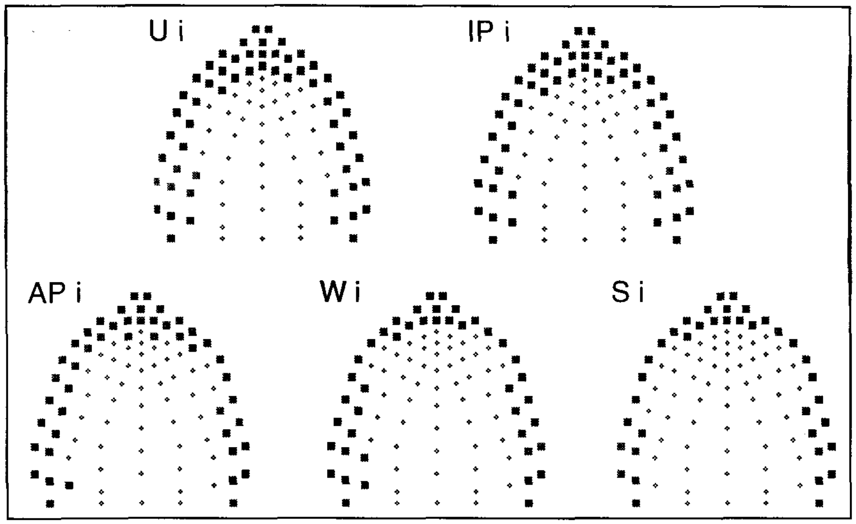


FIGURE 5: Patterns of linguopalatal contact for [n] depending on its Prosodic Position for one repetition of speaker 1F.

Discussion and Conclusion

In this study, we have shown that the magnitude of oral and nasal articulation of a consonant can be influenced by its position in the prosodic structure. This influence has been found in several prosodic levels in a single study and in a language not previously studied from this point of view: French. In general, we found that consonants in higher Prosodic Positions (that is consonants preceded by a strong prosodic boundary) have less nasal flow and more linguopalatal contact than consonants in lower Prosodic Positions.

The influence of Prosodic Position has been observed to apply to two remote articulatory subsystems: the tongue and the velum. However, the variation in the amount of nasal flow shows less distinction between Prosodic Positions than the amount of linguopalatal contact does. This result can be interpreted in two ways. (1) It could be that velum articulation distinguishes fewer positions than lingual articulation; a difference in degrees of freedom between the two articulators may prevent the velum from varying too much without endangering the velopharyngeal opening required to maintain nasality; for the tongue, on the contrary, greater variation in tongue height may be allowed behind the seal, as shown in the EPG profile in Fig. 5. (2) The difference in the sensitivity of the techniques (airflow is only an indirect measurement of velic articulation) could also explain the difference in the number of distinctions. These two hypotheses need to be tested in an experiment using the same technique of investigation for the 2 articulators (for example with an articulograph). In either case, the fact that Prosodic Position can influence these two articulators suggests that prosodic information is a high-level component of speech production control. As a consequence, we feel the need for a prosodic tier in a model of speech production in order to account for the articulatory characteristics due not only to prosodic prominence (stress or accent), but also to particular Prosodic Positions (or phrasing). Only differences in magnitude of articulatory movements have been discussed here, but Prosodic Position may also affect the timing of articulatory gestures as suggested in the coarticulation experiments presented in the introduction.

The patterns of variation of lingual and velic articulation may appear at first glance to be opposite: in higher Prosodic Positions, the contact of the tongue against the palate is increased for coronal consonants, whereas the velopharyngeal opening of nasal consonants seem to be reduced (as found by Krakow, 1989). These modifications seem to be contradictory if we looked at the effect of Prosodic Position as a "strengthening" of articulation. But from an acoustic or perceptual point of view, this increased magnitude of oral articulation and decreased magnitude of nasal articulation may contribute to the same goal: the increase of the consonantality, and more generally the salience, of the consonant in "strong" Prosodic Positions. Manuel [1991] suggested that the reduction of velopharyngeal opening (therefore nasal flow) for nasal consonants in word initial positions contributes to the reduction of the sonority of the consonant. To test this hypothesis, acoustic consequences of this variation in articulation need to be further examined.

The results presented here were observed for two consonants and two subjects. The influence of Prosodic Position found agrees with previous results, but it needs to be examined with several subjects. Previous experiments involving 3 to 4 subjects have shown that prosodic influence on both nasal airflow in French [Fougeron, 1996] and lingual articulation in English [Fougeron & Keating, 1996] vary depending on the subjects. In both studies, although almost every subject showed distinctions between the more extreme prosodic levels, the number and the nature of the Prosodic Positions distinguished by articulation was a function of the subject.

The results presented here also suggest that some of the variation found in articulatory studies may be controlled, explained or understood by looking at the prosody of the speech materials. Therefore, awareness of prosodic differences between utterances can turn some apparently random variation into predictable regularities of speech production. However, these prosodic differences may not be reflected in differences in punctuation. For example, we found a variation in linguopalatal contact between the consonants initial in an Accentual Phrase and those initial in an Intonational Phrase, and these positions are both marked by a comma. On the other hand, we did not find a distinction between Utterance- and Intonational Phrase- initial positions, and these two positions are marked by different punctuation: one is marked by a period, the other by a comma.

Moreover, articulatory variation may be a valuable source of information in understanding the prosodic organization of speech. The possible acoustic and perceptual relevance of these position-dependent articulatory variations needs to be defined and tested, but from a production point of view, variation in articulation may follow the prosodic hierarchy and somehow cue prosodic phrasing.

This research was supported in part by an allocation M.R.T. to the D.E.A. de Phonétique de Paris, and by the NSF grant SBR-9511118 to P. Keating.

REFERENCES:

Beckman, M., Edwards, J., and Fletcher, J. (**1992**). "Prosodic structure and tempo in a sonority model of articulatory dynamics," in *Papers in Laboratory Phonology II: Gesture, Segment, Prosody,* edited by G. Docherty and D. R. Ladd (Cambridge U. Press, Cambridge UK), Chap. 3, 68-86.

Benguerel, A.-P. (**1977**), "Velar coarticulation in French: a fiberscopic study". *Journal of Phonetics* 5:149-158

Byrd, D. M. (**1994**), *Articulatory timing in English consonant sequences.* PhD. diss. UCLA. *(UCLA Working Paper 86)*

DeJong, K. (**1995**), "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation." *J.A.S.A.* 97:491-504.

Fougeron, C. & P. Keating (**1995**) "Demarcating prosodic groups with articulation". *J.A.S.A.* 97-5, pt.2

Fougeron, C. (**1996**), "Variation de débit nasal en fonction de la position prosodique de [n] et [ã] en français". in Actes des XXIè Journées d'Etude sur la Parole, Avignon, France.

Fougeron, C. & P. Keating (**1996**) "Articulatory strengthening in prosodic domain-initial position". *UCLA Working Paper, this volume*

Fujimura, O. (**1977**), "Recent findings on articulatory processes" . In *Articulatory modeling and phonetics* (Carré, Descout & Wajskop eds.), GALF, 115-126.

Hardcastle, W.J. (**1985**), "Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences." *Speech Communication* 4:247-263.

Jun, S.-A. & C. Fougeron (**1995**), "The Accentual Phrase and the prosodic structure of French". *ICPhS 95*, vol. 2, 722-725

Jun, S.-A. (**1993**), *The phonetics and phonology of Korean prosody.* PhD diss., Ohio State Univ.

Krakow, R. A. (**1989**), *The articulatory organization of syllables: a kinematic analysis of labial and velic gestures.* PhD. diss.Yale Univ.

Krakow, R. A., Bell-Berti, F., & Wang, Q.E. (**1994**). "Supralaryngeal declination: evidence from the velum," in *Producing speech: A Festschrift for Katherine Safford Harris*, edited by F. Bell-Berti and L. Raphael (AIP Press, Woodbury NY), Chap. 23:333-353.

MacClean, M. (**1973**), "Forward coarticulation of velar movement at marked junctural boundaries", *J.S.H.R.* 16:286-296

Manuel, S. (**1991**), Some phonetic bases for the relative malleability of syllable-final vs. syllable-initial consonants. *ICPhS 91*, vol. 5:118-121

Pierrehumbert, J. & D. Talkin (**1992**), Lenition of /h/ and glottal stop. In *Lab.oratory Phonology II* (Docherty & Ladd eds.) Cambridge U. Press.

Vaissière, J. (**1988**), "Prediction of velum movement from phonological specifications". *Phonetica* 45:122-39.

Influence of microprosody on macroprosody: a case of phrase initial strengthening [1]

Sun-Ah Jun

1. Introduction

Cross-linguistically, it is widely accepted that f0 of a vowel onset is influenced by the preceding consonant type (i.e. the microprosody). For example, voiceless aspirated or tense consonants raise the f0 of the following vowel while voiced consonants and breathy consonants lower the f0 of the following vowel (Gandour 1974; Hombert 1978; Hombert et al. 1979). But, there are also studies showing no rise-fall dichotomy. Ohde (1984) measured nonsense syllables produced by American English speakers both in isolation and embedded in carrier phrases. He found that after voiced stops f0 was either level or slightly falling as in f0 after voiceless stops. Silverman (1984, from Silverman 1986) also found falling f0 after voiced as well as voiceless stops, though the fall was much shallower after voiced stops. From monotone German utterance, Kohler (1982, 1985) also found no falling f0 after voiced stops. Silverman (1986) claims that these differences of microprosody between voiced and voiceless consonants are an artifact of the experimental design: the intonational contour (i.e. the macroprosody) was not controlled for. The direction of f0 movement after release of a stop will show rise-fall dichotomy when it is in the middle of a rising intonation, but not in other intonational contexts. That is, the microprosody is influenced by the macroprosody (also see Steele 1986). In contrast, Kingston and Diehl (1994) present crosslinguistic data revealing that f0 perturbation pattern differs from language to language. Based on these results, they claim that the f0 perturbation may not be a low level phonetic phenomena.

Korean obstruents have also been shown to influence f0 of the following vowel (Kim 1965, Han & Weitzman 1970, Hardcastle 1973, Kagaya 1974): a higher f0 after voiceless aspirated stops and voiceless unaspirated (tense) stops but a lower f0 after voiceless breathy (lenis) stops. Interestingly, these segmentally induced f0 perturbations have been shown to play an important role in the intonation pattern of Korean (Jun 1989, 1993). An Intonational Phrase (=IP) of Korean can consist of more than one smaller phrase which is also tonally marked. This tonally marked small phrase is called an Accentual Phrase. When an Accentual Phrase initial consonant is either aspirated or tensed (both are [+stiff vocal cords] following Halle & Stevens (1971)), the phrase begins with a H tone and otherwise a L tone. That is, the tonal pattern of a non-IP-final Accentual Phrase in Seoul is either LHLH or HHLH and that of Chonnam is either LHL or HHL. (When an Accentual Phrase is in IP final, the Accentual Phrase final tone is preemptied by an IP boundary tone.) It was claimed that, in both dialects, in an Accentual Phrase longer than three syllables, the first and the second tone of an Accentual Phrase are associated with the first and the second syllable of the phrase, respectively. (For a detailed description of an Accentual Phrase, see Jun 1993.) But, it was not clear whether the segmentally triggered phrase initial H tone is phonetic (due to undershoot of L tone) or is phonologized, i.e. underlying. That is, since the second syllable of an Accentual Phrase is associated with a H tone, it would be possible that the phrase

---

[1] A shortened version of this paper was presented as a poster at the fifth meeting of the Laboratory Phonology conference at Evanston, Illinois, in July 1996. I thank J. Kingston for providing his ASA paper and all speakers who participated in this experiment and especially Cécile Fougeron for her help in recruiting French speakers.

initial H triggered by the microprosody may stay high phonetically due to the following H tone. If so, we expect the phrase initial H tone would persist for a shorter duration as in other languages when it is followed by a Low tone. On the other hand, if the H tone is underlying, we expect it will persist regardless of the following tone types, and it will furthermore be influenced by the constraints on the underlying tonal sequences.

The goal of this study is to determine the status, phonetic or phonological, of the accentual phrase initial H tone in Korean, and to examine how the pattern of microprosody in Korean differs from those in English and French in different prosodic positions and different intonational environments. English is chosen since most of the relevant studies are based on English data. French is chosen since it was shown in Jun & Fougeron (1995) that the Accentual Phrase of French has a similar tonal pattern (LHLH) to that of Seoul Korean.

2. Method

The target syllable CV was placed in four different prosodic/intonational positions shown in Table 1. In order to match prosodic conditions across languages, the target syllable was placed in a prosodically similar carrier sentence: for English, "This is a ___ note" or "This is ___"; for French, "Il a dit ___." (He said __); for Korean, "iɾɨmi ___ja." (The name is ___.). Data sets for each language are shown in the appendix. In Set 1, there were two Accentual Phrases for Korean and French, and the target CV was put at the beginning of the second Accentual Phrase (=AP) which has two syllables. For English, the target CV was in the nuclear pitch accented position. This prosodic condition is the same as that used in earlier studies (Lehiste & Peterson 1961, Lea 1973, Hombert 1978). In Set 2, the target CV was put after focused word, thus in the middle of an AP in French and Korean, and in a post-nuclear pitch accented position in English. Sentences in both sets were declaratives ending in a L boundary (%) tone. In Set 3, the target syllable was in the same position as in Set 1 but the sentences were interrogatives ending in a H% tone. Lastly, in Set 4, the target syllable was at the beginning of a four-syllable AP in French and Korean, and for English, it was an unstressed syllable immediately preceding a nuclear pitch accented syllable followed by two more syllables. (Similar data are tested in Silverman 1986.) Since both Korean and French have a H tone on the second syllable of an accentual phrase when the accentual phrase is longer than three syllables (i.e. Set 4), English data were designed to have a primary stress on the second syllable of trisyllabic target words. Since stress is lexical in English, all target CVs were embedded in a real word in English. Most vowels were [ɚ] except for those in Set 4 where the vowel was often [o]. For French, a set of real words containing the target syllable was included as a control data in addition to these four sets. These are listed in Set 0 of the Appendix. For Korean, 10 out of the 13 words in each set except Set 4 were real words. But all other nonsense words satisfy Korean phonotactics, and thus were easily pronounceable by native speakers of Korean. For both French and Korean nonsense words, the vowel was [i].

English consonants examined were /b, d, g, v, z, m, l, p, t, k, f, s, h/ and French consonants were /b, d, g, v, z, m, p, t, k, f, s/, and Korean had 9 lenis, tense, aspirated stops plus /m, l, s, h/. Four speakers of each language, Parisian French, Seoul Korean, and Californian English, participated in the experiment. Each subject repeated sentences 7 to 10 times. The utterances were digitized and analyzed using Entropic Research Laboratory's XWAVES speech analysis software. f0 was measured at vowel onset, and 20, 40, 60, 80, and 100 ms into the vowel.

Table 1. Four prosodic conditions

set #	sentence format	prosodic conditions
Set 1	{σσσ}{CV σ}.	AP initial in French & Korean In nuclear pitch accented syllable in English Followed by L%
Set 2	{σσσ CV σ}.	AP medial in French & Korean In post-nuclear pitch accented syll in Eng. Followed by L%
Set 3	{σσσ}{CV σ}?	AP initial in French & Korean In nuclear pitch accented syll in English Followed by H%
Set 4	{σσσ}{CV σσ(σ)}.	AP initial in French & Korean Before nuclear pitch accented syll in Eng. Followed by H phrase or H* tone

3. Results and Discussion

Since the f0 pattern of sonorant consonants in English and French was the same as that after voiced obstruents, sonorant consonants were grouped together to form a voiced consonant group to compare with the group of voiceless consonants for all languages. Mean f0 after each voiced vs. voiceless consonants from English and French is shown in Figure 1 and 3, respectively. For Korean, since the f0 pattern of sonorant consonants was the same as that after lenis stops, sonorant consonants were grouped together with lenis stops. The mean f0 of lenis stops and sonorants vs. other voiceless consonants from Korean is shown in Figure 4. The X-axis in these figures refers to the time from the vowel onset until 100ms in every 20 ms. In each figure, Set 1 data is shown in the upper left graph, Set 2 in the upper right graph, Set 3 in the lower left graph, and Set 4 in the lower right graph. For all languages, the place where the target CV syllable appears is indicated as a square box. The error bar in each graph represents one standard deviation.

As shown in Figure 1, f0 values after voiceless and voiced consonants in English merged sooner than those in previous studies (e.g. Hombert 1978): 20-40 ms rather than 80-100ms. This pattern is similar to the pattern shown in Kingston and Diehl (1994, p. 433). Furthermore, f0 after voiced stops did not rise under the same intonation patterns as those used in early studies. That is, the target CV in English Set 1 and Set 4 is the same prosodic condition as in Hombert (1978) but the f0 after voiced consonants was either level or falling as in f0 after voiceless consonants. This is the same pattern found in Ohde (1984) where nonsense words were used. The f0 difference between voiced vs. voiceless consonants was even smaller in Set 2 and Set 3, indicating that the consonant perturbation pattern is indeed influenced by the overall intonational contours (Kohler 1982, Steele 1986, and Silverman 1986). When the target CV is in the middle of a falling intonation, (i.e. Set 2) or before rising (Set 4), f0 perturbation pattern was very similar between voiced and voiceless consonants. A similar pattern is also shown in Kingston & Diehl (1994, p. 434).

For French data, the f0 perturbation pattern after voiced vs. voiceless consonants in real words (Set 1) was in general very similar to that in nonsense words (Set 0) as shown in Figure 2. Overall, f0 showed more variation in real words. Also, f0 after voiceless consonants in real words tends to be more separated from f0 after voiced consonants for most speakers than they are in nonsense words. This may be caused by the different types of segments following the word initial consonant in real words. But, in both cases, f0 after voiced consonants was either level or rising and f0 after voiceless consonants was falling as claimed in previous studies. This f0 separation between voiced and voiceless consonants in real words was not significant from 20 to 40ms after consonant onset, and this timing was the same in nonsense words. This verifies that as far as a f0 pattern is concerned, using nonsense words for French data was not different from using real words. f0 values after voiced and voiceless consonants in French also merge around 20-60ms after vowel onset, but unlike in English, f0 after voiced consonants is rising except for that in Set 2. This may be due to the fact that French voiced stops are more often fully voiced than English voiced stops. Again, the f0 pattern after voiced and voiceless consonants is less different in Set 2, showing the influence of macroprosody. Basically, the f0 pattern in Set 2 was similar in all three languages.

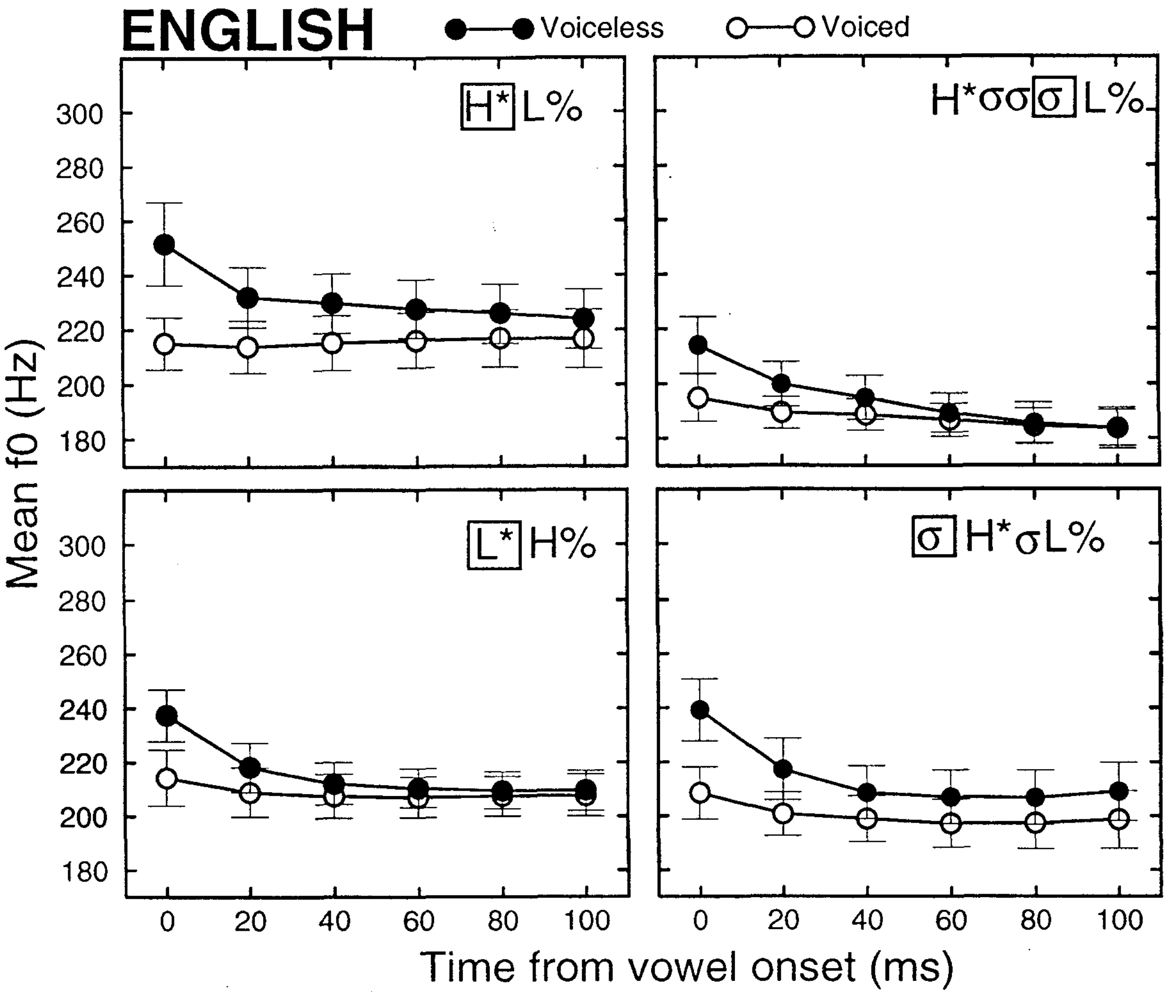


Figure 1. Mean f0 (in Hz) pattern of English voiceless (filled circle) vs. voiced (empty circle) consonants

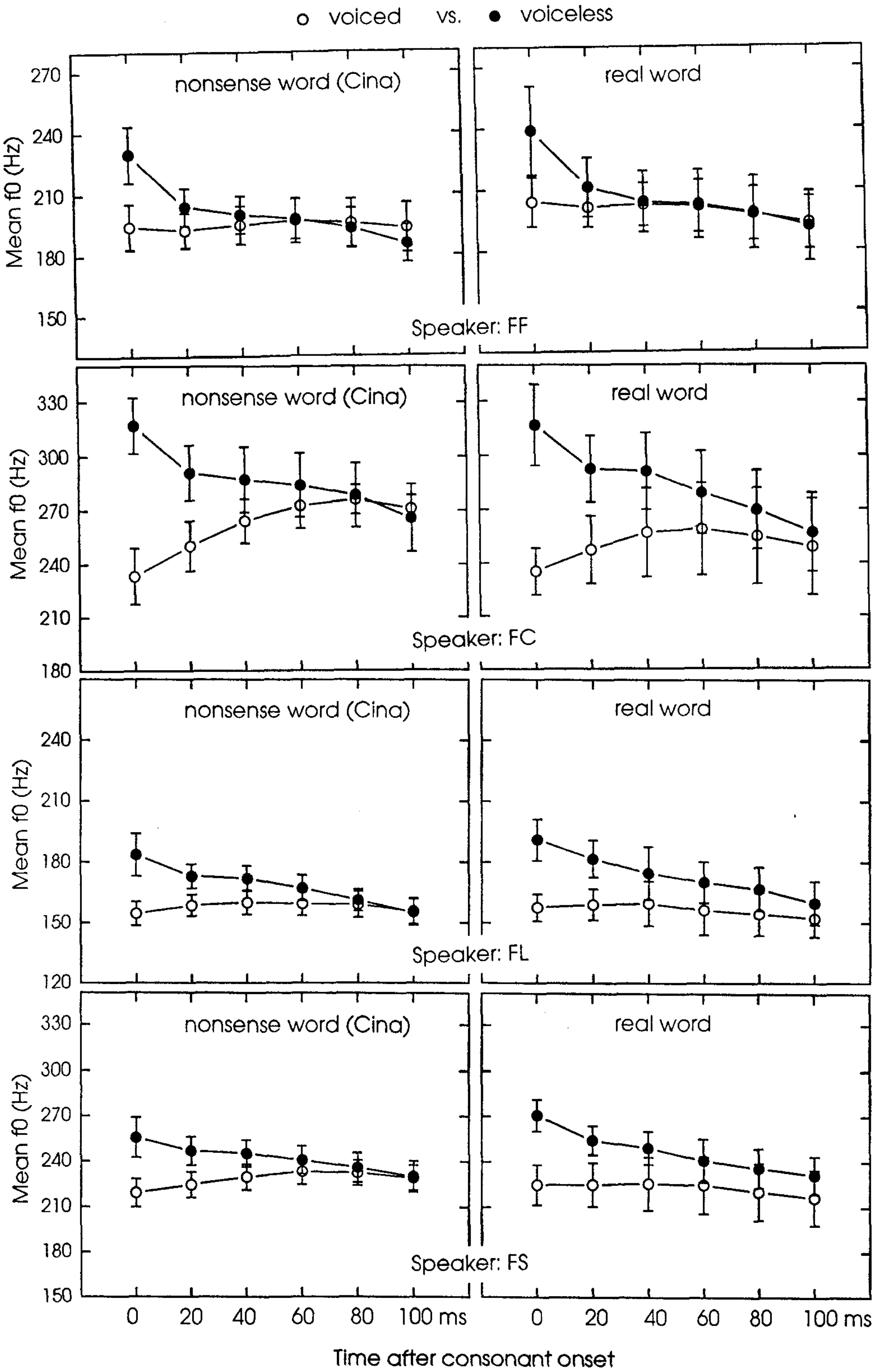


Figure 2. Mean f0 (in Hz) pattern of French Set 0 (real words) and Set 1 (nonsense words) produced by four French speakers

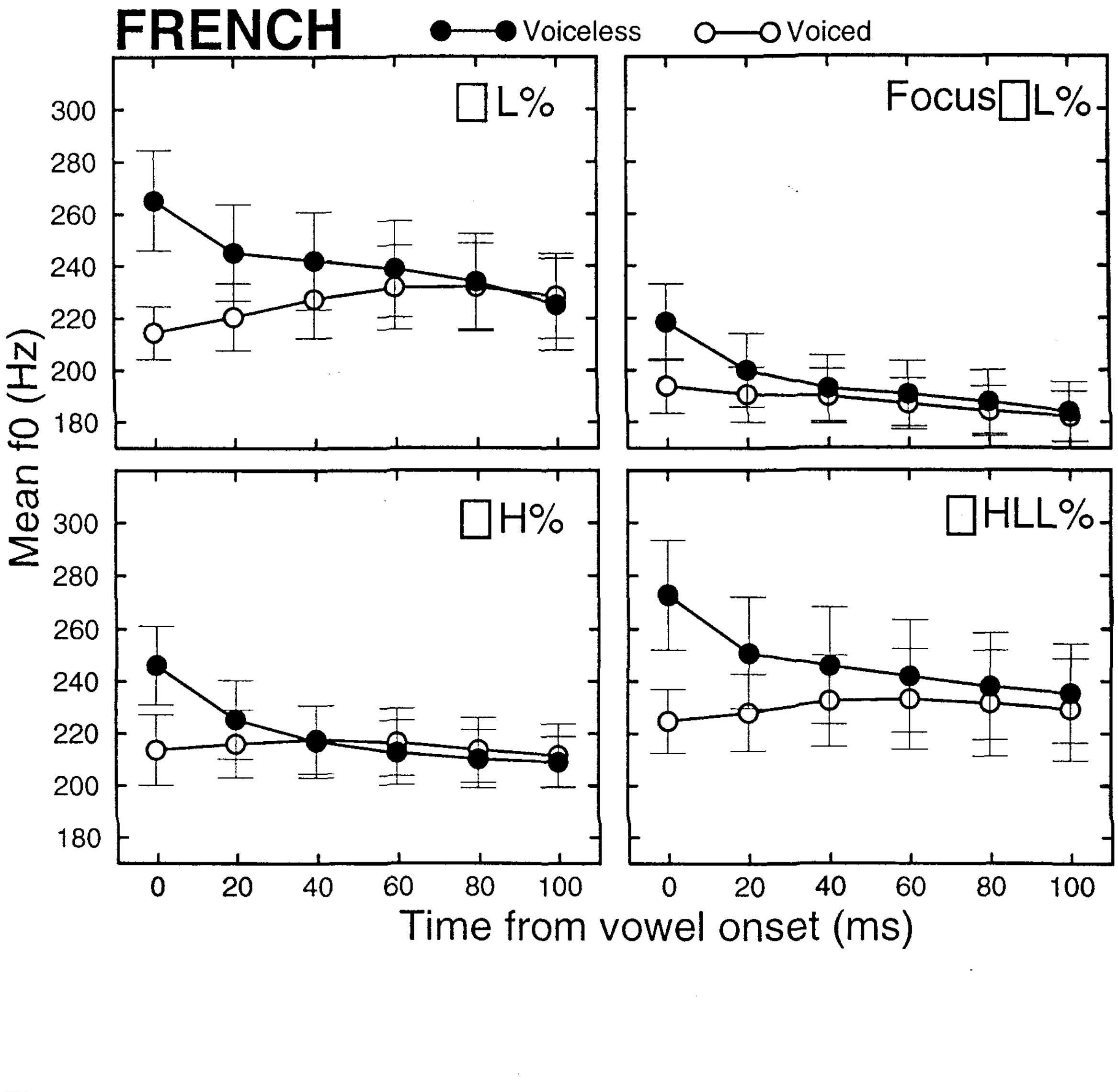


Figure 3. Mean f0 (in Hz) pattern of French voiceless (filled circle) vs. voiced (empty circle) consonants

The f0 pattern after Korean consonants (Figure 4) was very different from English and French, except for that in Set 2. For Korean, as noted in Jun (1993), f0 after aspirated and tense stops was significantly higher (in average 50-80 Hz) than that after lenis stops and sonorant consonants, and these f0 differences persisted until the end of the vowel. This f0 difference was far greater than that found in English or French, suggesting that the f0 perturbation type in Korean is not the same as that in other languages. If the phrase initial H tone results from L tone undershoot due to the following H tone, we would expect a similar pattern of f0 values both in English and French, especially in Set 4 where the phrase initial syllable is followed by H* in English and by phrasal H in French. But as shown in Figure 1 and 3, the f0 values differ only at phrase initial position and the difference does not persist longer than 40-60 ms into the vowel. In addition, the phrase initial high f0 triggered by [+stiff vocal cords] consonants in Korean remained high regardless of the following tone types and showed a tonal interaction with the following tones, suggesting the phrase initial H tone in Korean is not due to phonetic undershoot but is part of the underlying representation of intonation. For example, when the phrase initial consonant has a [+stiff vocal cords] feature, the phrase initial syllable had a H tone even before the L boundary tone (%) of a declarative sentence (Set 1 in Figure 4). At the same time, when the syllable was followed by an interrogative marker which is H% by default, the boundary tone became either LH% or upstepped H% or HL%, as shown in Figure 5.

Further evidence for an underlying phrase initial H tone was found in EMG data from a Chonnam speaker (the author). From EMG data (Thyroarytenoid (Vocalis) muscle and Posterior Cricoarytenoid muscle) collected for vowel devoicing (Jun and Beckman in preparation), we found that the Vocalis muscle was highly active at a vowel onset after tense consonants, but also active in parallel with a high f0 contour. Representative examples are shown in Figure 6. In the first example, the last Accentual Phrase begins with an aspirated stop [p^h], and pitch tracks show a high f0 value. At the same time, the Vocalis muscle also shows a high level of activity, about 80 ms before the timing of high f0 in pitch track, as marked by an arrow. (EMG activity trace is 80 ms advanced compared to acoustic waveform and airpressure data.) The second example shows when the last Accentual Phrase begins with a lenis stop /t/. f0 values in the pitch tracks are low during this syllable and the corresponding Vocalis activity is also low, lower than the following peak which is aligned with the phrasal H tone of the second syllable of the phrase. The Vocalis muscle trace (first window) still shows a high level of activity after [+stiff vocal cords] consonants even when the vowel /i/ was devoiced as indicated by an arrow in Figure 7. The waveform (fourth window) shows frication of /s/ but no periodic movement before stop closure for /k'/ (tense velar stop) and the corresponding intraoral air pressure trace (second window) shows no dip during the vowel portion, indicating a complete devoicing of vowel [i].

However, when the consonant was in the middle of an Accentual Phrase, the Vocalis activity was changed following the f0 pattern. The same sentences shown in Figure 6 but with focus on the word preceding the target word is shown in Figure 8. In Korean, a focused word always initiates an Accentual Phrase and all the following unfocused words are dephrased showing a L tone (Jun 1993). Thus, the target syllable, [p^hu], in the upper graph, now has a L tone and the corresponding Vocalis muscle activity is also lowered (indicated by an arrow) compared to the case where the consonant was at the beginning of an Accentual Phrase (Figure 6). For the same reason, the target syllable, [tu], in the lower graph shows a L tone and low Vocalis muscle activity (indicated by an arrow), thus neutralizing with the phrase medial aspirated syllable in the upper graph.

In sum, f0 and EMG pattern show that phonologization of pitch perturbation in such a way it affects Korean intonation is only limited to an Accentual Phrase initial position. In other words, in Korean, microprosodically triggered tone is phonologized only phrase initially. This limitation can be interpreted as a phrase initial strengthening. Crosslinguistically, production data have been shown that acoustic duration as well as lingual contact duration of EPG is longer phrase initially than phrase medially, indicating that segments at the phrase initial position are strengthened (Jun 1993 for Korean; Hsu and Jun 1996 for Taiwanese; Fougeron 1996 for French; Fougeron and Keating 1996 for English; Gordon 1996 for Estonian). That is, by producing a High or Low tone on the vowel following the phrase initial consonant, the perception of three consonant types, i.e. the three-way contrast either in the laryngeal feature or VOT/closure duration, is enhanced in the phrase initial position (Stevens and Keyser 1989), but not in the phrase medial position. However, since there are only two tonal contrasts in phrase initial position, H for [+stiff vocal cords] consonants and L for [-stiff vocal cords] consonants, this would imply that what is enhanced is not the distinction among three consonant types, but between lenis versus the other two tense consonant types. This way of strengthening seems to match well with the acoustic/articulatory characteristics of these consonants. Among three types of consonants, lenis consonants have in general the weakest acoustic/articulatory property. That is, aspirated stops have the longest VOT and largest oral airflow, and tense stops have the longest closure duration, the strongest burst energy, and the strongest oral airpressure, but lenis stops have the medium or weak value in all these phonetic dimensions. Thus, the perceptual saliency of the lenis stop itself is weaker than other stops. Thus, by producing a Low tone following the phrase initial lenis stop but a High tone following the other two types of stops, the perceptual saliency of the lenis stop would increase and, so, the contrast is enhanced.

Employing f0 perturbation to enhance a phonological contrast has been claimed in Kingston (1986). He argues that f0 perturbation is phonologized to enhance a distinctive feature of [voice]. Thus, languages where [voice] is not a distinctive feature as in Tamil, f0 perturbation does not need to be phonologized "because speakers are not trying to transmit a contrast between [+voice] and [-voice] stops, but between short and long ones". However, [voice] is not a distinctive feature in Korean, either, (all obstruents are voiceless) but f0 perturbation is still phonologized to enhance the contrast between lenis vs. other two types of consonant. This suggests that f0 perturbation in Korean is used to enhance [tense] feature which was claimed by Kim (1965) to distinguish [-tense] (= lenis stops) consonants from [+tense] (=aspirated and tense stops) consonants. What is interesting in Korean is that enhancing [tense] feature through the phonologization of f0 perturbation depends on the prosodic position of the consonant: it is limited to the phrase initial position.

This interpretation is supported by the perception experiments reported in Cho (1996) and Han (1996). Cho (1996) found that native Korean subjects were able to perceive (67%) the prevocalic consonants based on f0 and intensity of the following vowel. Interestingly, subjects perceived lenis stops and tense stops more accurately (71~83% and 65%~86%, respectively) than aspirated stops (42~57%). Furthermore, aspirated consonants were more often misperceived as tense consonants (52%) than as lenis consonants (10%), and tense consonants were more often perceived as aspirated consonants (26%) than as lenis consonants (12%). This pattern of error matches the pattern of f0 following the consonants. In this experiment, the perception token was from the first syllable (CV) of a two-syllable word produced in a citation form, thus a phrase initial CV. Han (1996, pp. 174-179), on the other hand, tested if the f0 pattern found in production data is also utilized in the perception of the stops by Korean speakers. She created synthetic stimuli of tense and lenis stops followed by a vowel, and varied the onset f0 values continuously from

122Hz to 158Hz for each stops. Then she asked Korean subjects to identify the CV stimulus either as a lenis or tense stop. The result showed that "the subjects perceived the stimuli systematically, with lower f0 giving rise to lenis stops and higher f0 giving rise to tense stops", and the category boundary was 134 Hz when the consonant part of the stimulus was from a lenis stop, and 143Hz when the consonant was from a tense stop. This indicates that there is an inherent quality of lenis vs. tense consonants, and f0 strongly enhances the perception of the consonant. Here again, the CV was also at word initial and phrase initial position. These two perception experiments show that the tonal difference produced after stops is perceptually salient when stops are in the phrase initial position. It would be very interesting to examine the case when the perception token is from an accentual phrase medial position.

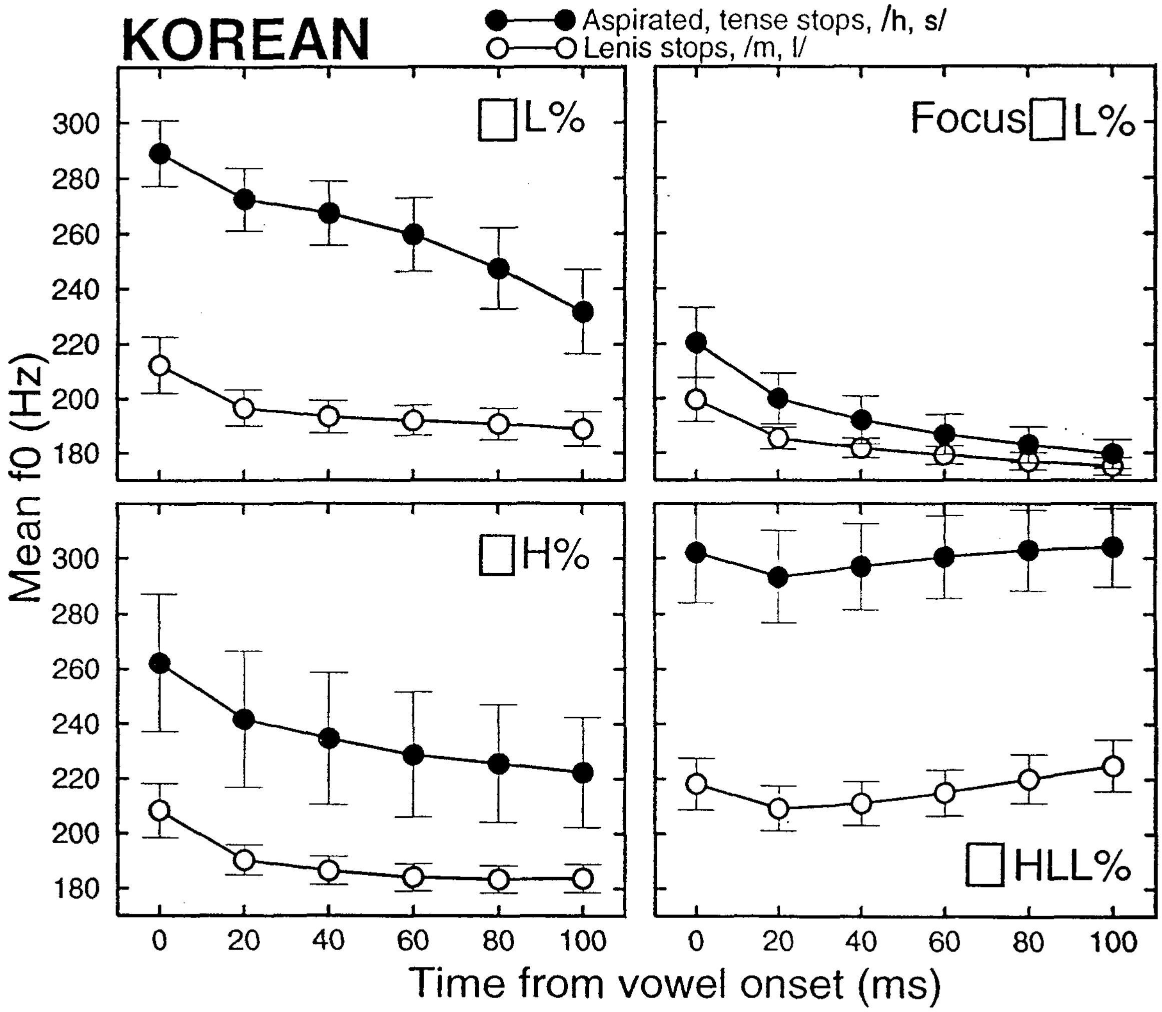


Figure 4. Mean f0 (in Hz) pattern of Korean: aspirated and tense stops, and /h, s/ group is shown as a filled circle, and lenis stops and /l, m/ group is shown as an empty circle.

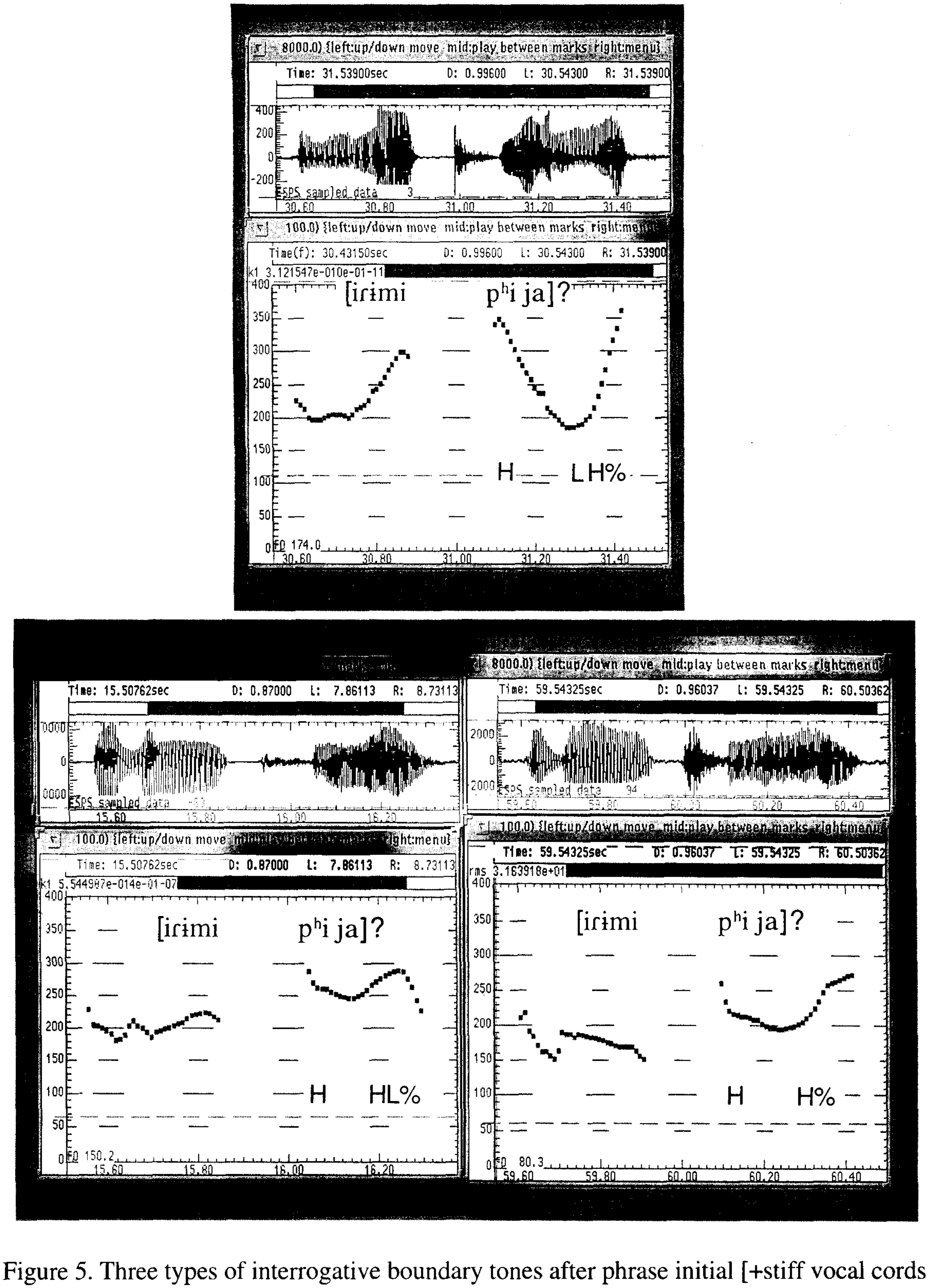


Figure 5. Three types of interrogative boundary tones after phrase initial [+stiff vocal cords] consonant, /p^h/.

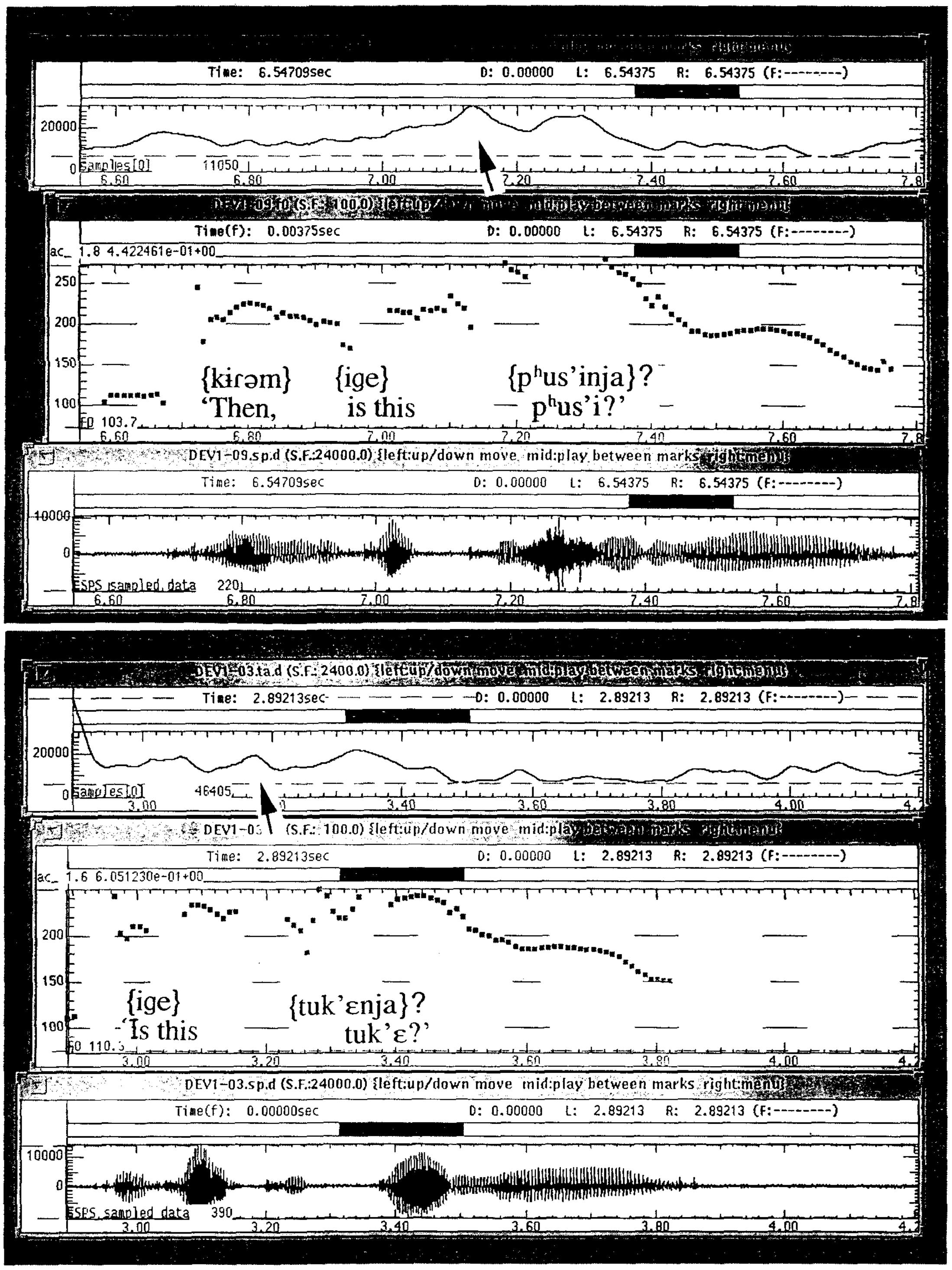


Figure 6. Examples of Vocalis muscle activity, f0 tracks, and waveforms when the Accentual Phrase initial segment is [+stiff vocal cords], /pʰ/, (above) and [-stiff vocal cords] consonant, /t/ (below).

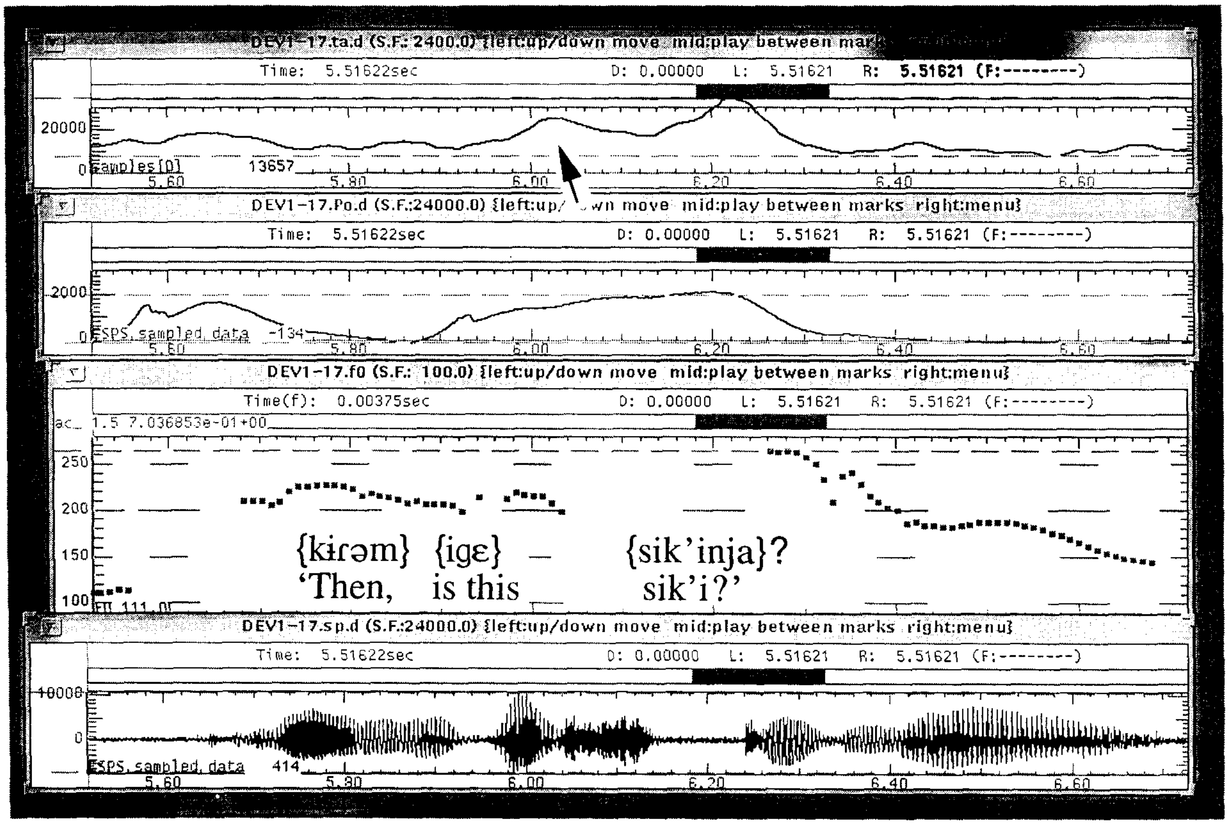


Figure 7. An example of Vocalis muscle activity (first window), intraoral pressure (second window), f0 tracks (third window), and waveforms (fourth window) when a vowel /i/ is devoiced after an Accentual Phrase initial [+stiff vocal cords] consonant.

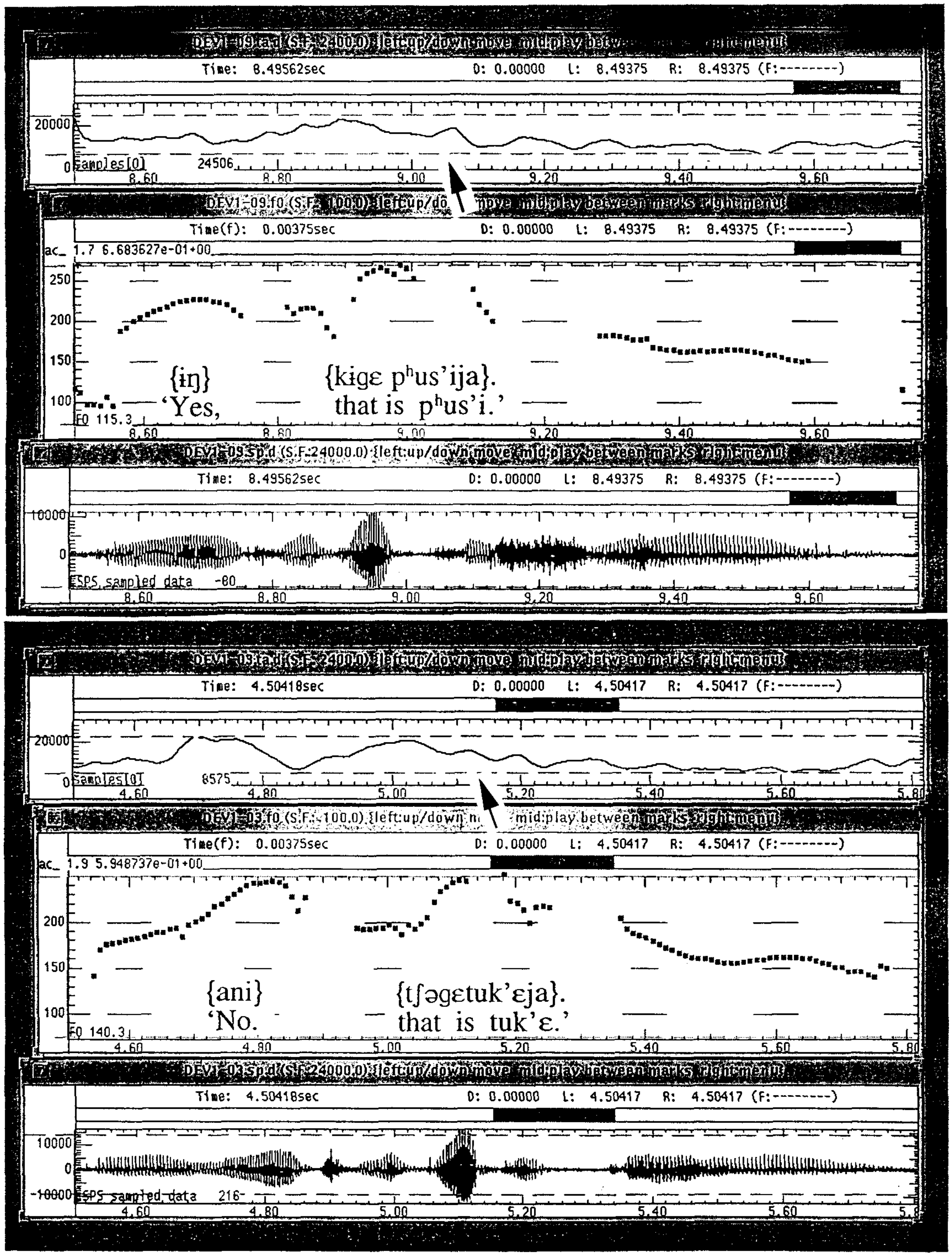


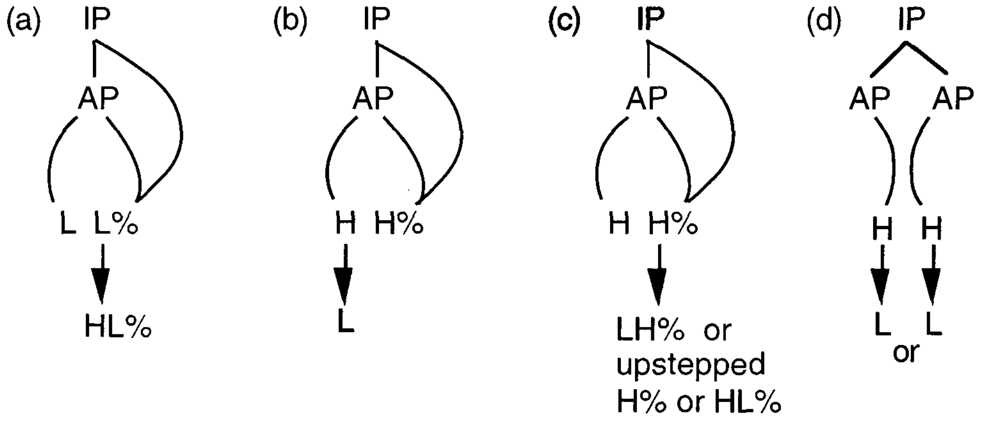
Figure 8. Examples of Vocalis muscle activity, f0 tracks, and waveforms when the target syllable, /pʰ/ (upper) and /t/ (lower) is in the middle of an Accentual Phrase.

Next, in addition to phrase initial strengthening due to tonal contrast triggered by a consonant, it was found that there is a constraint on tonal sequence in Korean which is also sensitive to a phrase boundary. I call this constraint the 'see-saw effect'.

(1) The 'see-saw' effect: A sequence of two identical tones tends to become different from each other if they are boundary tones belonging to a different prosodic level (AP or IP) or belonging to a different prosodic group ({AP}{AP}).

For example, an accentual phrase initial H tone followed by L% was not changed, but an accentual phrase initial L tone followed by L% became L and HL%. This is schematically represented in (2a) below and an example pitch track is shown in Figure 9(a). In the same way, an accentual phrase initial L tone followed by H% sequence was not changed, but an accentual phrase initial H tone followed by H% was changed so that they are not identical. That is, as shown in Figure 5, H% after the phrase initial H became one of the three types: LH% or upstepped H% or upstepped HL% . A schematic representation of the tonal change is shown in (2c) below. Alternatively, the phrase initial H can become L before H% as schematized in (2b) below. An example pitch track is shown in Figure 9(b). All these changes have something in common: one of the two adjacent tones changes to sound different from each other, either categorically (from L to H or vice versa) or phonetically (from H to upstepped H).

(2) schematic representation of the 'see-saw' effect



Cases shown in (2 a, b, c) show interaction between tones belonging to a different prosodic level. But, tones belonging to the same prosodic level but belonging to a different prosodic grouping also behave similarly. For example, when an Accentual Phrase initial H is preceded by an Accentual Phrase final H, one of the H often becomes a L tone. However, when the phrase initial H is also followed by H% which marks an Intonational Phrase boundary, the phrase initial H often becomes L. This explains why there is a large variation in Set 3 of Figure 4 above, larger than the variations shown in Set 1 or Set 4. In Set 4, where the phrase initial H is preceded by a phrase final H but followed by a phrase medial H which does not mark any phrase boundary, the phrase initial H was less likely to change into a L tone. Rather, the preceding phrase final H tone often became L tone. A schematic representation of this change is shown in (2d) and an example pitch track is shown in Figure 9(c). In all cases, the tones were changed to avoid a sequence of the same tone. But, when both tones are within the same prosodic group as in the Accentual Phrase initial HH sequence, no tones were changed. So, it seems that the 'see-saw' effect applies when both tones are marking a boundary of a prosodic group. I believe there is a perceptual motivation for the 'see-saw' effect: namely, identical tones are perceptually hard to distinguish in sequence. Thus, when they mark prosodically important boundaries, they change in order to be perceived better.

4. Conclusion

In this paper, it was shown that unlike English and French, the phrase initial H tone in Korean is triggered by segmental perturbation, and that this tone is phonologized in Korean. This segmental perturbation in Korean triggers an underlying tonal difference only in phrase initial position, thus, supporting the phrase initial strengthening phenomena found in the production data from several languages. Finally, a sequence of two identical boundary tones belonging to different prosodic levels or different prosodic grouping tends to be modified so that they differ from each other: the 'see-saw' effect.

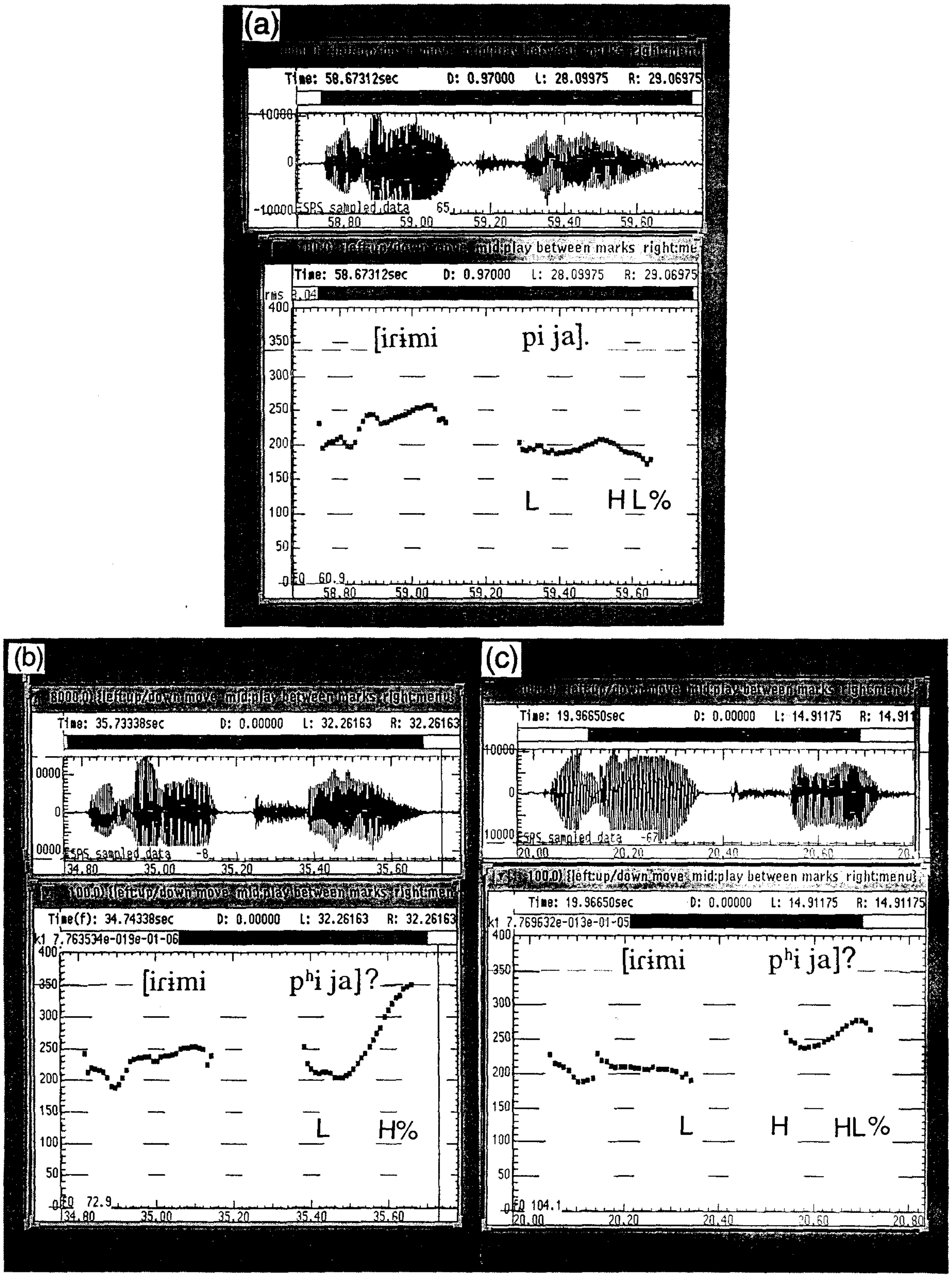


Figure 9. Tone interaction patterns showing the 'see-saw' effect. (a) L% after L becomes HL%, (b) H before H% becomes L, and (c) phrase final H becomes L before phrase initial H.

References

Cho, T. (1996) *Vowel correlates to consonant phonation: an acoustic-perceptual study of Korean obstruents.* MA thesis. The University of Texas at Arlington.

Fougeron, C (1996) "Variation de débit nasal en fonction de la position prosodique de [n] et [ɑ̃] en Français", *XXIes Journeés d'étude sur la Parole.*

Fougeron, C. & P. Keating (1995) "Demarcating prosodic groups with articulation", a paper presented at the 129th meeting of the ASA. Washington, DC.

Fougeron, C. & P. Keating (1996) "The Influence of Prosodic Position on Velic and Lingual Articulation in French: Evidence from EPG and Airflow data", in the *Proceedings of the 1st ESCA Tutorial and Research Workshop on Speech Production Modeling.* pp. 93-96.

Gandour, J. (1974) "Consonant types and tone in Siamese", *J. of Phonetics* 2.

Gordon, M. (1996) "Nasal duration and amplitude as a function of stress and prosodic phrasing in Estonian", a paper presented at the 131st meeting of ASA. Indianapolis, IN.

Halle, M & K. Stevens (1971) "A note on laryngeal features", *MIT Quarterly Progress Report* 101: 198-212. Research Laboratory of Electronics, Cambridge, MA.

Han, J.-I. (1996) *The Phonetics and Phonology of "Tense" and "Plain" consonants in Korean.* Diss. Cornell University.

Han, M. and R. Weitzman (1970) "Acoustic features of Korean /P, T, K/, /p, t, k/, and /p^h, t^h, k^h/", *Phonetica* 22:112-128

Hardcastle, W. (1973) "Some observations on the tense-lax distinction in initial stops in Korean", *Journal of Phonetics* 1: 263-272

Hombert, J.-M. (1978) "Consonant types, vowel quality, and tone", in *Tone: A linguistic survey.* ed. by V. Fromkin. 77-111.

Hombert, J.-M., J. Ohala, & E. William (1979) "Phonetic explanations for the development of tones", *Language* 55:37-58.

Hsu, C. and S.-A. Jun (1996) "Is Tone Sandhi Group part of the Prosodic Hierarchy in Taiwanese?", a paper which will be presented at the 132nd meeting of the ASA, Hawaii.

Jun, S.-A. (1989) " "The Accentual Pattern and Prosody of the Chonnam Dialect of Korean," in S. Kuno et al. (eds.) *Harvard Studies in Korean Linguistics III.*, Harvard Univ., Cambridge, MA

Jun, S.-A. (1993) *The Phonetics and Phonology of Korean Prosody.* Diss. The Ohio State Univ.

Jun, S.-A. & M. Beckman (in preparation) "Articulatory study of Korean vowel devoicing" ms.

Jun, S.-A. & C. Fougeron (1995) "The Accentual Phrase and the Prosodic structure of French", in the *Proceedings in XIIIth ICPhS*, Vol. 2:722-725.

Kagaya R. (1974) "A Fiberscopic and Acoustic Study of the Korean Stops, Affricates, and Fricatives," *Journal of Phonetics* 2, 161-180.

Kim, C.-W. (1965) "On the autonomy of the tensity feature in stop classification (with special reference to Korean stops)", *Word* 21: 339-359

Kingston, J. (1986) "Are f0 differences after stops accidental or deliberate?", a paper presented at the 111th meeting of the Acoustical Society of America, Spring 1986.

Kingston, J. & Diehl (1994) "Phonetic Knowledge", *Language* 70(3)

Kohler, K. (1982) "F0 in the production of lenis and fortis plosives", *Phonetica* 39.

Kohler (1985) "f0 in the perception of lenis and fortis plosives", *JASA* 78: 21-32

Lehiste, I. & G. Peterson (1961) "Some basic considerations in the analysis of intonation", *JASA* 33: 419-425.

Lea, W. (1973) "Segmental and suprasegmental influences on f0 contours," in Hyman (ed.) *Consonant types and tones. S. Calif. Occas. Papers in Ling.* 1:15-70.

Ohde, R. (1984) "Fundamental frequency as an acoustic correlates of stop consonant voicing", *JASA* 75: 224-230.

Silverman, K. (1984) “f0 perturbations as a function of voicing of prevocalic and postvocalic stops and fricatives, and of syllable stress”, in Lawrence, *Proc. Autumn Conf. Inst. Acoustics*, Windermere, Vol 6, 445-452.

Silverman, K. (1986) “F0 segmental cues depend on intonation: The case of the rise after voiced stops”, *Phonetica* 43: 776-91

Steele, S. (1986) “Interaction of vowel F0 and prosody”, *Phonetica* 43: 92-105

Stevens, K. and S. Keyser (1989) “Primary features and their enhancement in consonants,” *Language* 65: 81-106.

Appendix (Words in italics are focused.)

English Data

Set 1. 0. This is a “learn” note.
1. This is a “perm” note.
2. This is a “turn” note.
3. This is a “kern” note.
4. This is a “fern” note.
5. This is a “seal” note.
6. This is a “burn” note.
7. This is a “meal” note.
8. This is a “durn” note.
9. This is a “girl” note.
10. This is a “Verne” note.
11. This is a “zeal” note.
12. This is a “heel” note.

Set 1. 0. *This* is a “learn” note.
1. *This* is a “perm” note.
2. *This* is a “turn” note.
3. *This* is a “kern” note.
4. *This* is a “fern” note.
5. *This* is a “seal” note.
6. *This* is a “burn” note.
7. *This* is a “meal” note.
8. *This* is a “durn” note.
9. *This* is a “girl” note.
10. *This* is a “Verne” note.
11. *This* is a “zeal” note.
12. *This* is a “heel” note.

Set 3 0. Is this a “learn” note?
1. Is this a “perm” note?
2. Is this a “turn” note?
3. Is this a “kern” note?
4. Is this a “fern” note?
5. Is this a “seal” note?
6. Is this a “burn” note?
7. Is this a “meal” note?
8. Is this a “durn” note?
9. Is this a “girl” note?
10. Is this a “Verne” note?
11. Is this a “zeal” note?
12. Is this a “heel” note?

Set 4. 0. This is “logician”.
1. This is “pomaceous”.
2. This is “tornado”.
3. This is “Koranic”.
4. This is “foramen”.
5. This is “Somalia”.
6. This is “Bohemian”.
7. This is “momentum”.
8. This is “donation”.
9. This is “Gorgonian”.
10. This is “voracious”.
11. This is “zoology” [zoʊɑlədʒi].
12. This is “homology”.

French data:

Set 0. 1. il a dit piler.
2. il a dit tisane.
3. il a dit kiné.
4. il a dit fila.
5. il a dit sigma.
6. il a dit bilan.
7. il a dit disons.
8. il a dit guider.
9. il a dit villa.
10. il a dit zigzag.
11. il a dit miner.

Set 1.
1. il a dit pina.
2. il a dit tina.
3. il a dit kina.
4. il a dit fina.
5. il a dit sina.
6. il a dit bina.
7. il a dit dina.
8. il a dit guina.
9. il a dit vina.
10. il a dit zina.
11. il a dit mina.

Set 2.
1. il a *dit* pina.
2. il a *dit* tina.
3. il a *dit* kina.
4. il a *dit* fina.
5. il a *dit* sina.
6. il a *dit* bina.
7. il a *dit* dina.
8. il a *dit* guina.
9. il a *dit* vina.
10. il a *dit* zina.
11. il a *dit* mina.

Set 3.
1. At-il dit pina?
2. At-il dit tina?
3. At-il dit kina?
4. At-il dit fina?
5. At-il dit sina?
6. At-il dit bina?
7. At-il dit dina?
8. At-il dit guina?
9. At-il dit vina?
10. At-il dit zina?
11. At-il dit mina?

Set 4.
1. il a dit pinanamou.
2. il a dit tinanamou.
3. il a dit kinanamou.
4. il a dit finanamou.
5. il a dit sinanamou.
6. il a dit binanamou.
7. il a dit dinanamou.
8. il a dit guinanamou.
9. il a dit vinanamou.
10. il a dit zinanamou.
11. il a dit minanamou.

<u>Korean data</u>:

[iɾɨm-i] 'a name-NOM.', [li-ja] 'li-a declaratives or an interrogative ending of be'
=> "The name is 'li'".

Set 1.

1. 이름이 리야.	[iɾɨmi lija]
2. 이름이 비야.	[iɾɨmi pija]
3. 이름이 피야.	[iɾɨmi p^hija]
4. 이름이 삐야.	[iɾɨmi p'ija]
5. 이름이 시야.	[iɾɨmi sija]
6. 이름이 히야.	[iɾɨmi hija]
7. 이름이 미야.	[iɾɨmi mija]
8. 이름이 디야.	[iɾɨmi tija]
9. 이름이 티야.	[iɾɨmi t^hija]
10. 이름이 띠야.	[iɾɨmi t'ija]
11. 이름이 기야.	[iɾɨmi kija]
12. 이름이 키야.	[iɾɨmi k^hija]
13. 이름이 끼야.	[iɾɨmi k'ija]

Set 2.

1. *이름이* 리야.	[iɾɨmi lija]
2. *이름이* 비야.	[iɾɨmi pija]
3. *이름이* 피야.	[iɾɨmi p^hija]
4. *이름이* 삐야.	[iɾɨmi p'ija]
5. *이름이* 시야.	[iɾɨmi sija]
6. *이름이* 히야.	[iɾɨmi hija]
7. *이름이* 미야.	[iɾɨmi mija]
8. *이름이* 디야.	[iɾɨmi tija]
9. *이름이* 티야.	[iɾɨmi t^hija]
10. *이름이* 띠야.	[iɾɨmi t'ija]
11. *이름이* 기야.	[iɾɨmi kija]
12. *이름이* 키야.	[iɾɨmi k^hija]
13. *이름이* 끼야.	[iɾɨmi k'ija]

Set 3.
1. 이름이 리야? [irɨmilija]
2. 이름이 비야? [irɨmipija]
3. 이름이 피야? [irɨmi p^{h}ija]
4. 이름이 삐야? [irɨmi p'ija]
5. 이름이 시야? [irɨmisija]
6. 이름이 히야? [irɨmihija]
7. 이름이 미야? [irɨmimija]
8. 이름이 디야? [irɨmitija]
9. 이름이 티야? [irɨmi t^{h}ija]
10. 이름이 띠야? [irɨmi t'ija]
11. 이름이 기야? [irɨmikija]
12. 이름이 키야? [irɨmi k^{h}ija]
13. 이름이 끼야? [irɨmi k'ija]

Set 4.
1. 이름이 리아나래. [irɨmilianare]
2. 이름이 비아나래. [irɨmipianare]
3. 이름이 피아나래. [irɨmi p^{h}ianare]
4. 이름이 삐아나래. [irɨmi p'ianare)
5. 이름이 시아나래. [irɨmisianare]
6. 이름이 히아나래. [irɨmihianare]
7. 이름이 미아나래. [irɨmimianare]
8. 이름이 디아나래. [irɨmitianare]
9. 이름이 티아나래. [irɨmi t^{h}ianare]
10. 이름이 띠아나래. [irɨmi t'ianare]
11. 이름이 기아나래. [irɨmikianare]
12. 이름이 키아나래. [irɨmi k^{h}ianare]
13. 이름이 끼아나래. [irɨmi k'ianare]

Phrase-Final Lengthening and Stress-Timed Shortening in the Speech of Native Speakers and Japanese Learners of English*

Motoko Ueyama
(ueyama@humnet.ucla.edu)

1. Introduction

1.1. Problem

Difficulties in acquiring the temporal organization of English have been observed among native speakers of Japanese (Mochizuki-Sudo & Kiritani, 1991). Those difficulties appear to be caused by distinctive differences between Japanese and English in multiple aspects. In addition to the inventory size of segment and syllable types, the two languages greatly differ in prosodic realization, including temporal organization. The typological categorization of timing systems is rooted in the idea that temporal organization is based on some unit of timing, and Japanese is said to be mora-timed, whereas English is stress-timed (Dauer, 1983). English timing is based on *stress accent* (i.e. on a local prosodic prominence that is lexically defined). One of the phonetic correlates of stress accent is duration, which is alternately stretched and reduced, interacting with the other two correlates, namely, fundamental frequency and amplitude. The unit of English timing is the stress foot, that is, a string containing a stress accent followed by zero or more unstressed syllables (Beckman and Edwards, 1990). On the other hand, the unit of Japanese timing is the mora (a syllabification unit): the duration of each mora is equal, abstracting away from the phrase-final lengthening effect that we will discuss later. The prosodic distinctions of Japanese is mainly conveyed by fundamental frequency (Beckman, 1986). These differences in the manipulation of the duration cue with respect to the other prosodic cues suggest that native speakers of Japanese will have considerable difficulty in learning English timing.

1.2. Two Hypothesized Effects on English Timing Patterns

Two crucial concepts concerning English temporal organization are *phrase-final lengthening* and *stress-timed shortening*. The phrase-final lengthening effect is usually defined as the lengthening of a rhyme (nucleus and coda) occurring before the boundary between prosodic constituents, roughly reflecting syntactic boundary strength (cf. Wightman et al., 1992 for English; Beckman & Pierrehumbert, 1986; Kaiki *et al,* 1992 for Japanese).

The stressed-timed shortening effect is an indication of the tendency toward "isochronous spacing of prosodically strong syllables; a stressed syllable in a polysyllabic word or stress foot is compressed in order to make the overall duration of its word or stress foot closer to that of a contrasting monosyllable (Beckman & Edwards 1990; 152)." Intervals between two stressed syllables are called "interstress intervals (ISIs)." The size of the stress foot is frequently used to denote the size of ISIs. The most prominent element in the stress foot is called the head of the foot. In the theory of isochrony, intervals between two heads are expected to be equalized. Consequently, the durations of stressed syllables will be shortened as the number of unstressed

---

* This is a short version of my MA thesis, titled *Phrase-Final Lenthening and Stress-Timed Shortening Effects in Native Speakers and Japanese Learners of English* (1996) with additional data and analysis. For further discussion, please refer to the original thesis. This paper will be published in the proceedings of the Fourth International Conference on Spoken Language Processing.

syllables between the two heads increases. This effect should characterize stress-timed languages.

1.3. Purpose of the Study

This study is intended to investigate the acquisition of phrase-final lengthening and stress-timed shortening by Japanese learners of English in comparison with native speakers of English. As Beckman and Edwards (1990) point out, in the experimental design, it is crucial to prevent the two effects from being confounded with each other. The production experiment of Mochizuki-Sudo and Kiritani (1991) is subject to this criticism. For the purpose of this study, the two effects will be strictly distinguished in experimental design. The durational patterns of native Japanese speakers learning English will be described and analyzed with respect to the transfer of the Japanese timing system to the timing system of their English speech .

2. Experiment

2.1. Subjects

The set of speakers included one control group and two experimental groups; each group consisted of four speakers (two males and two females). The control group (NS group) consisted of native speakers of American English. The first experimental group (the AJ group) consisted of advanced Japanese learners of English, and the second experimental group (the BJ group) consisted of beginning Japanese learners of English. In order to equalize proficiency across speakers in the same experimental group, the Japanese participants were selected using several criteria: years of residence in the U.S. (3 years for the AJ; none for the BJ), age range (25-31 for the AJ; 22-25 for the BJ), language background, and previous learning experience.

2.2. *Procedure*

The subjects' assigned task was to read a list of randomized sentences. Each sentence was read twelve times, and the first and last two repetitions were omitted in the subsequent measurement of phonetic durations. The performances were recorded with high-quality equipment in the soundproof room of the UCLA Phonetics Laboratory for the AJ group and in the recording booth of the Doosisya Women's College in Japan for the BJ group.

The recorded data were converted from analog to digital at a 10 kHz sampling rate, and analyzed using Kay Elemetrics' Computerized Speech Laboratory (CSL) hardware and software. Since final lengthening mainly affects phrase-final syllables (Wightman, et al., 1992), and the relevant syllables had no consonantal coda, I measured the duration of stressed and unstressed vowels in both pre- and post-boundary positions for each prosodic level. All the measurements were carried on the waveform analysis, and wide-band spectrographs were additionally inspected in the cases of difficult segmentation.

2.3. *Materials*

The data sets for the production experiment are shown in (1-A, B, C). The two stressed vowels determining the size of ISI are indicated by acute accent marks; the analyzed stressed and unstressed vowels are denoted by bold-face and underline, respectively. The three sentence sets (1-A, B, C) test the three degrees of boundary strength predicted by the theory of prosodic hierarchy (see Hayes 1989 for a review): boundary strength is weakest for the set (1-A) and strongest for the set (1-C).

(1)

A. Boundary separates **members of a compound**
  a. I met L**ée**] Mótt last week.
  b. I met **Lée**] DeMótt last week.
  c. I met L**é**da] Mótt last week.
  d. I met L**é**da] DeMótt last week.

B. Boundary follows the **end of a phonological phrase**.
  a. My p**á**] zíps along in traffic.
  b. My p**á**] posítions himself well in traffic.
  c. My p**á**pa] zíps along in traffic.
  d. My p**á**pa] posítions himself well in traffic.

C. Boundary follows the **end of an intonational phrase**.
  a. My p**á**], góat lover though he is, would never buy one.
  b. My p**á**], pagóda lover though he is, would never buy one.
  c. My p**á**pa], góat lover though he is, would never buy one.
  d. My p**á**pa], pagóda lover though he is, would never buy one.

For each set, the pair (a, b) strictly tests for two different degrees of stress-timed shortening (ISI = 0 vs. ISI = 1), holding phrase-final lengthening constant (i.e. putting a boundary in the same position relevant with respect to the pre-boundary noun). In the same way, the pair (c, d) examines the difference between ISI = 1 and ISI = 2. The pair (b, c) strictly tests for phrase-final lengthening, holding the size of interstress interval constant (ISI = 1). The structure of the experimental design is summarized in **Table 1**. These pairs were compared in two-tailed t-tests.

Table 1. Sentence pairs tested for phrase-final lengthening and stress-timed shortening.

test for	informative pairs	expected final lengthening	ISI size
phrase-final lengthening	(b) vs. (c) e.g. p**á**] posítions vs. p**á**pa] zíps	*yes* for stressed V in (b) and unstressed V in (c); *no* for unstressed V in (b) and stressed V in (c)	constant (ISI =1)
stress-timed shortening	(a) vs. (b) e.g. p**á**] zíps vs. p**á**] posítions	constant (*yes* for stressed V)	ISI = o vs. 1
	(c) vs. (d) e.g. p**á**pa] zíps vs. pápa]posítions	constant (*no* for stressed V)	ISI = 1 vs. 2

3. Results & Discussion

Vowel durations in each condition were statistically analyzed by using ANOVA and two-tailed t-tests. The effects of the two prosodic phenomena under consideration were examined separately.

3.1. Similar Patterns in the Three Groups

The results of the NS group reliably showed a major phrase-final lengthening effect before both phonological-phrase and intonation-phrase boundaries, but word-final lengthening effects varied across and within speakers. The data from a representative speaker, NS1, are shown in **Figure 1**, and the statistical results are summarized in **Table 2** (The shaded cells indicate p-values above the critical value ($\alpha = 0.05$). IF, PF, and WF are abbreviated for Intonational-Phrase Final, Phonological-Phrase Final, and Word Final boundaries. Similar patterns were observed in the data of all the four AJ speakers and the four BJ speakers.

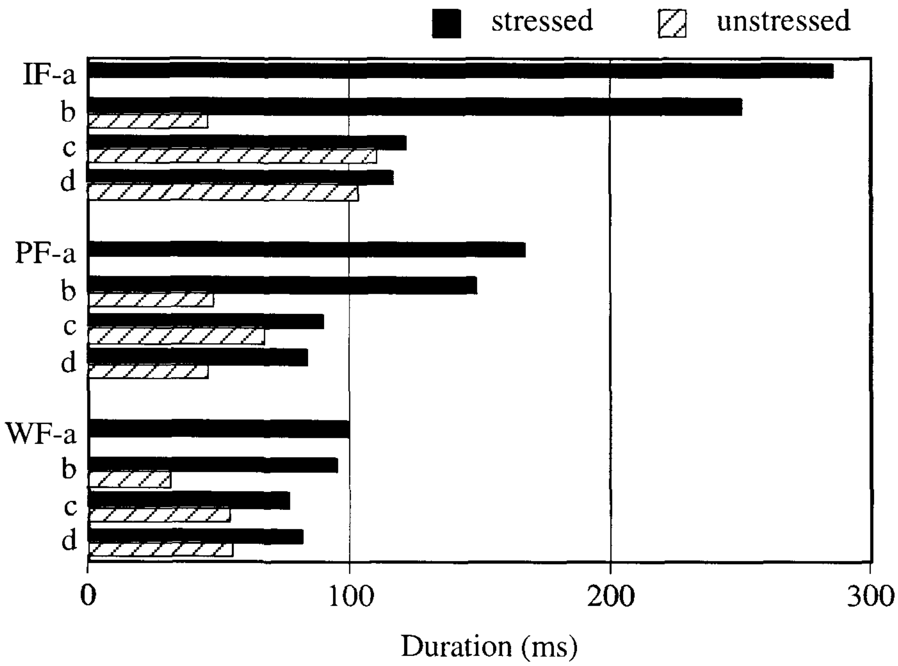


Figure 1. Mean durations of stressed and unstressed vowels of NS1.

Table 2. Results of two-tailed t-tests in p-values for the data of NS1

		WF	PF	IF
Stressed V				
phrase-final lengthening	(b) vs. (c)	.059	p<.001	p<.001
stress-timed shortening	(a) vs. (b)	.697	.172	.022
	(c) vs. (d)	.118	.230	.435
Unstressed V				
phrase-final lengthening	(b) vs. (c)	.059	p<.001	p<.001

3.2. Major Effects in Japanese Learners' Speech

The production data of the three groups showed considerable phrase-final lengthening and no stress-timed shortening, and these patterns were statistically confirmed. What factor differentiates the durational patterns of the three groups, if any? Further statistical analyses show the following two major peculiarities in the speech of Japanese learners.

First, less advanced Japanese learners make less durational contrast between lexically stressed and unstressed vowels than native English speakers. I assume this to be due to the negative transfer of Japanese "durational immalleability" (i.e. the fact that there is no durational contrast between pitch-accented and unaccented moras). The strong correlation between the speaker groups and the degree of durational contrast between stressed and unstressed vowels is shown by plotting the mean ratio of stressed V to an unstressed V adjacent to a prosodic boundary (i.e. not in a final lengthening context) for the set (1-C). The results are shown in **Figure 2**:

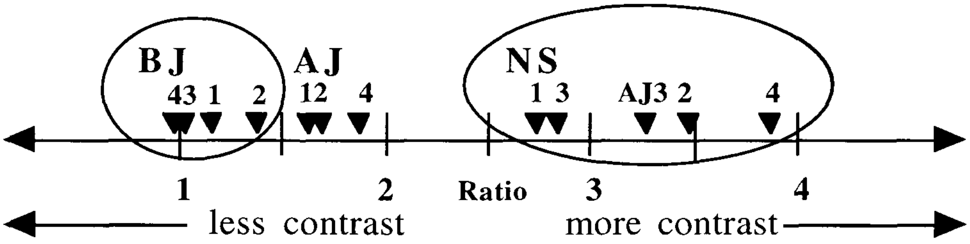


Figure 2. Ratio of mean durations. Stressed V in (c) pápa], góat were divided by unstressed V in (b) pá], pagóda for the (1-C) set.

Second, less advanced learners made less hierarchical differentiation of boundary strength. Additional statistics was conducted in order to see how the different groups differentiate the three degrees of boundary strength by duration. For each sentence set, the means of stressed vowels in the environment (b) were compared with the means of stressed vowels in the environment (c): e.g. (b) *pá], pagóda* vs. (c) *pápa], góat* for the (1-C) set. This comparison allows us to measure the differentiation of boundary strength and the degree of final lengthening by holding the ISI size constant (ISI = 1). The three sentence sets designed for this study reflect three degrees of boundary strength (WF < PF < IF). This order is expected to be systematically marked by the degree of final lengthening. The mean ratio of stressed vowels in the environment (b) to unstressed vowels in the environment (c) is compared across the three groups in **Figure 3**. The NS group clearly distinguishes the three degrees of boundary strength in the hypothesized order WF < PP < IF. Overall, the native speakers have more final lengthening for stressed syllables before stronger boundaries. This pattern is followed by the AJ group except AJ1. The patterns of the BJ group show more variations across the speakers. BJ1 seems to show almost native-like differentiation of the three levels, whereas the other three speakers do not. BJ2 makes no differentiation, and BJ3 and BJ4 make minor differentiation.

However, the similar pattern in ratio does not imply a similar pattern in actual magnitude of lengthening. **Figure 4** shows raw values of vowel durations in the three prosodic positions for the three speaker groups. In the environment (b), all native English speakers significantly increase the actual duration proportionally (WF < PF < IF). The same hierarchical differentiation is made in the environment (c) as well. This pattern is shown by only two Japanese speakers

(AJ2 and AJ4). The other six Japanese speakers tend to over-apply final lengthening before smaller boundaries. This results in a greater magnitude of lengthening before smaller boundaries and less hierarchical differentiation of boundary strength.

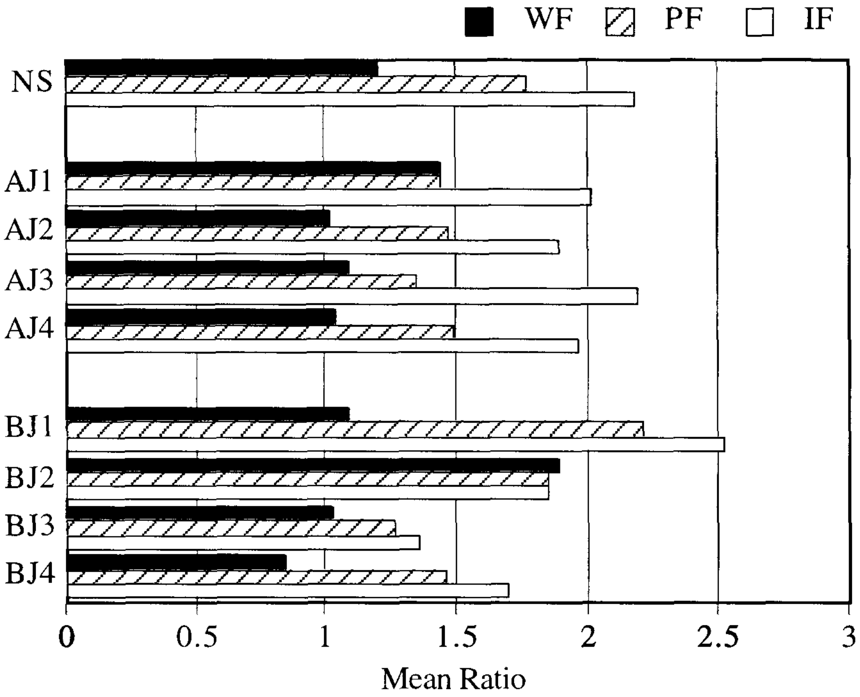


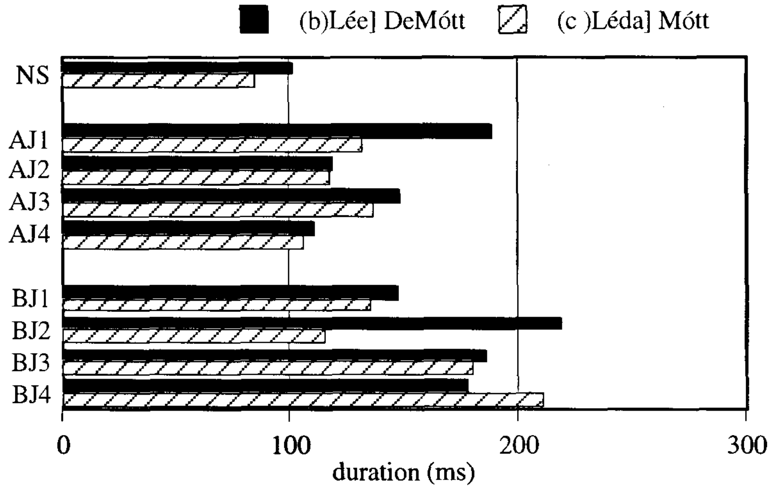
Figure 3. Mean ratio of stressed vowels in (b) to stressed vowels in (c).

4. Conclusion

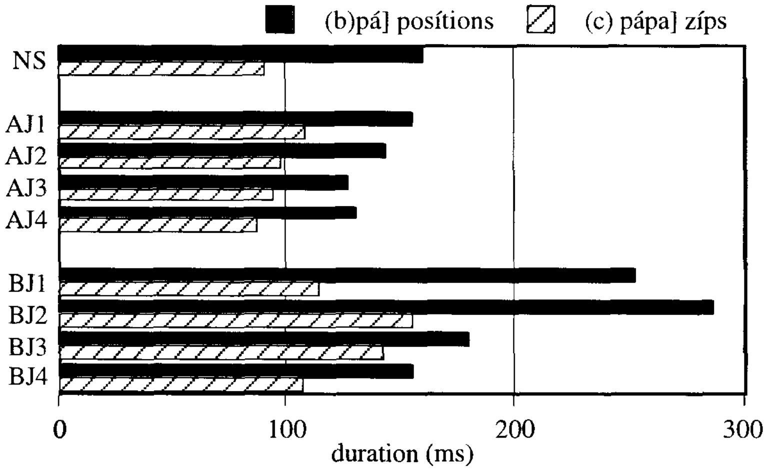
The detailed comparison of the three speaker groups demonstrates that Japanese timing patterns are influential in two ways. First, there is a primary negative transfer of mora immaleability, which is a Japanese specific feature. This constitutes a major difficulty in acquiring durational contrasts between lexically stressed and unstressed syllables in English. Second, there is a possible positive transfer of Japanese prepausal lengthening in larger boundaries, which may be a language-universal phenomenon.

To further explore the general mechanism of timing development in second language phonetics, we need cross-linguistic investigation of other cases, such as the English speech of second language learners whose first language is syllable-based (e.g. French) and the Japanese speech of second language learners whose first language is stress-timed (e.g. English).

A. Word Final



B. Phonological-Phrase Final



C. Intonational-Phrase Final

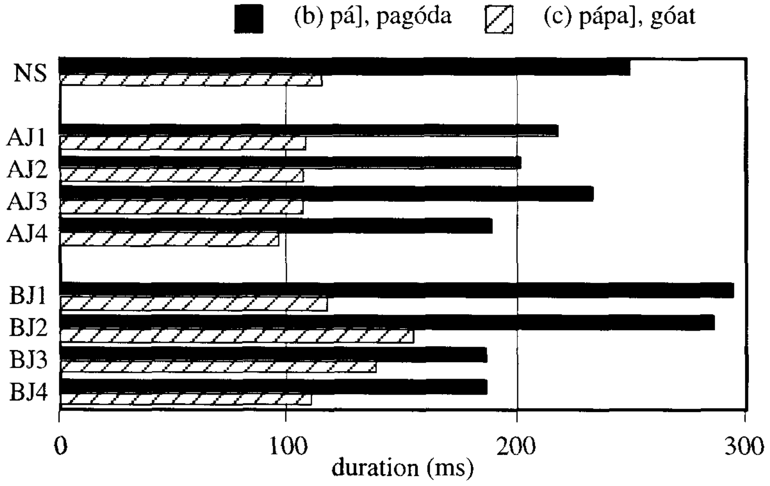


Figure 4: Mean durations of pre-boundary stressed vowels in the environments (b) and (c)

Acknowledgments

Grateful thanks are extended to all the speakers who participated in the experiments for this study. I would like to thank Sun-Ah Jun and Bruce Hayes for their continued advice and encouragement. I am also thankful for technical advice and support to Patricia Keating and my fellow students in the UCLA Phonetics Laboratory. Especially, I thank Marco Baroni and Joe Buffington, who volunteered to proof-read the manuscript. Also, I am grateful for the comments given on a preliminary version (presented in the 1995 Fall meeting of ASA) by Mary Beckman, Hiroya Fujisaki and Yoh'ichi Tohkura. The responsibility for any errors is solely mine.

References

Beckman, M. E. (1986). *Stress and Non-stress Accent*. Dordrecht: Foris Publication.

Beckman, M. E., & Edwards, J. (1990). Lengthening and shortening and the nature of prosodic constituency. In *Laboratory Phonology I: Between the Grammar and the Physics of Speech* (J. Kingston and M. E. Beckman, Editors), pp. 152-178. Cambridge: Cambridge University Press.

Beckman, M. E. & Pierrehumbert, J. (1986). Intonational structure in Japanese and English, *Phonology Yearbook,* **3**, 255-309.

Dauer, R. (1983). Stress-Timed and Syllable-timing reanalyzed, *J of Phonetics,* **11**, 51-62.

Hayes, B. P. (1989). The Prosodic Hierarchy in Meter. In *Phonetics and Phonology, Volume 1: Rhythm and Meter* (P. Kiparsky & G. Youmans, Editors), pp. 201-260. San Diego: Academic Press.

Kaiki, N. et al. (1992). Linguistic properties in the Control of Segmental Duration for Speech Synthesis. In *Talking Machines: Theories, Models, and Designs* (G. Bailly, et al., Editors). Elsvier Science Publishers B. V.

Mochizuki-Sudo, M. & Kiritani, S. (1991). Production and Perception of Stress-related Durational Patterns in Japanese Learners of English, *Journal of Phonetics,* **19**, 231-248.

Wightman, C. W., et al. (1992). Segmental Durations in the Vicinity of Prosodic Phrase Boundaries, *Journal of the Acoustical Society of America,* **91**, 1707-1717.

THE DEVOICING OF /Z/ IN AMERICAN ENGLISH: EFFECTS OF LOCAL AND PROSODIC CONTEXT

Caroline L. Smith*

1. INTRODUCTION

Although it is well-known that speakers of English often produce phonologically voiced stop consonants such as /b/ with little or no vibration of the vocal folds (Lisker & Abramson 1964, 1967), the extent to which voiced fricatives are also produced without vocal fold vibration has been investigated much less. This article examines the production of a voiced fricative, /z/, using simultaneous acoustic, aerodynamic and electroglottographic data. Using the term "voiced" to refer to /b/, /z/, and similar sounds implies that they belong to a category of sounds whose representation includes the conditions favorable for vocal fold vibration: an articulatory setting for vocal fold adduction and an aerodynamic setting for a pressure drop across the glottis. Studies by Haggard (1978), Veatch (1989), Docherty (1992), and Stevens, Blumstein, Glicksman, Burton & Kurowski (1992) have shown that speakers of both American and British English often produce phonologically voiced fricatives as wholly or as partially devoiced, so that either there is no vibration of the vocal folds or it does not last as long as the frication noise created at the oral constriction. If devoicing occurs during the fricative, there must be an adjustment to the state of the glottis, the aerodynamic conditions in the vocal tract, or the degree of constriction in the oral cavity. This article investigates how any of these adjustments, alone or in combination, might lead to devoicing, and proposes a representation in terms of the Articulatory Phonology model of Browman & Goldstein (1986) for the temporal coordination of these adjustments.

1.1 Proposed explanations for the tendency of voiced fricatives to be devoiced

Probably the most widespread explanation for why speakers devoice phonologically voiced fricatives is that simultaneous voicing and frication are difficult to produce (Ohala 1983). (This hypothesis proposes that voiced fricatives make more stringent requirements on the speech production mechanism than most other sounds.) In common with other voiced sounds, voiced fricatives require that subglottal pressure be higher than oral pressure in order to maintain vibration of the vocal folds. In common with voiceless fricatives, high airflow through the oral constriction is necessary in order to create the characteristic noise of the fricative. These two requirements compete in a voiced fricative. Oral pressure must be kept relatively low (compared to subglottal pressure) to maintain voicing, but the narrow oral constriction will tend to increase pressure in the mouth. Although oral pressure needs to be kept low, if too much air is vented from the mouth, the airflow through the oral constriction would not generate turbulence, as is necessary for a fricative. Calculations based on the cross-sectional areas of glottal and oral constrictions and pressures in different parts of the vocal tract show that in the absence of maneuvers explicitly directed at prolonging glottal vibration, it may not last the entire duration of the oral constriction in a voiced fricative (Stevens et al 1992).

This difficulty hypothesis is a production-based explanation for the rarity of voiced fricatives in the world's languages and their potential for devoicing in languages where they do occur (Ohala 1983). An alternative hypothesis (Balise & Diehl 1994) suggests that voicing is perceptually disfavored for fricatives, particularly sibilants, because it diminishes their characteristic high

---

* Now at Department of Linguistics, University of Ottawa.

intensity noise. From a perceptual point of view, the need for distinctiveness among the sounds in a language might seem likely to discourage devoicing in English, since each of the voiced fricatives has a voiceless counterpart. Viewed as a conflict between ease of production and ease of perception, devoicing may be an example of sounds that are easier to produce (devoiced fricatives) winning out over those that are more easily identified (fully voiced fricatives). However, experiments on the perception of fricatives have shown that listeners may use frequency and duration information in the signal, not just the presence of glottal vibration, in deciding whether to categorize a fricative as voiced or voiceless (Raphael 1972; Soli 1982; Baum & Blumstein 1987; Jongman 1989; Stevens et al 1992).

1.2 Production of voiced fricatives

If voiced fricatives are indeed difficult to produce, as Ohala (1983) suggested, it might be expected that the goal of any modification to their production would be to simplify whatever makes them so difficult, thus changing the acoustic output of the vocal tract. In the case of voiced fricatives, the difficulty seems to lie primarily in the rather specific aerodynamic requirements for the combination of vocal fold vibration and frication. Successful production of a fricative requires high intraoral pressure to maintain the flow of air through the oral constriction. However, if the oral pressure is allowed to rise too much, it may become too high compared to the subglottal pressure, which needs to be higher than the oral pressure in order for the vocal folds to vibrate. If the airflow from the lungs diminishes, the subglottal pressure could also diminish too much, again leading to the cessation of vocal fold vibration. On the other hand, if the oral constriction is wider, more air will be vented from the mouth and the oral pressure will be low. This situation would facilitate maintenance of a pressure drop across the glottis that is adequate for voicing, but might not maintain sufficient oral pressure to create the turbulent airflow that is necessary for frication noise (Ohala 1983). The result would be a voiced approximant rather than a fricative. At least for English, there do not seem to be any reports in the literature of voiced fricatives being simplified into voiced approximants, while simplification in the form of a loss of voicing is quite common (Haggard 1978, Docherty 1992). This asymmetry suggests that speakers are simplifying voiced fricatives by allowing oral pressure to rise and/or subglottal pressure to fall, not by widening the oral constriction and allowing oral pressure to fall. This study will attempt to distinguish among these hypotheses by comparing the production of voiced or devoiced /z/ and voiceless /s/.

The /s/ and /z/ notation is used here to indicate two "segments", or complexes of gestures, that are lexically contrastive in English (and many other languages). As suggested by Byrd (1994), the notion of segment could be interpreted as a group of articulatory gestures (in the sense of Browman & Goldstein 1986) occurring within a single word with the temporal phasing of the gestures defined with a narrow tolerance for variation. /s/ includes both a lingual constriction and a glottal opening; /z/ has a lingual constriction gesture that is probably identical to that of /s/ (Scully 1971), but with glottal adduction rather than glottal opening. One difference between /s/ and /z/ is that while vocal fold activity for /z/ can range from continuous vibration to complete absence of vibration, /s/ is consistently produced with little or no vibration (Docherty 1992).

1.3 Amount and likelihood of devoicing

A phonologically voiced fricative is said to be devoiced if it lacks vocal fold vibration during all or part of its duration, which is defined as the time during which there is strong aperiodic noise. In a fricative that is partially devoiced, most often the voicing is present during the first part of the fricative and then ceases before the end of the noise. Since the percentage of fricative duration that is devoiced will vary in different productions, comparisons can be made of the amount of devoicing that occurs in different environments. Stevens et al (1992) report larger amounts of devoicing in voiced fricatives followed in a cluster by a voiceless fricative than in those that are followed by a voiced fricative (difference of 13-29 ms). Similar results were obtained by Docherty (1992), who found on average 30 ms less voicing in /z/ following /ʃ/ than in intervocalic /z/.

Comparing fricatives produced in different contexts also make it possible to study the likelihood of devoicing in these different contexts. This variable has received more attention in previous studies than the amount of devoicing. One reason for this emphasis may be because likelihood of devoicing in different contexts can be studied by making a binary determination (voicing/devoicing) for each individual token, whereas comparison of the amount of devoicing in different tokens requires that the entire time course of each token be examined. Previous studies that have investigated the frequency of occurrence of devoicing have found substantial variation among different phonological contexts. For British English, Haggard (1978) found that fricatives following a voiced stop were devoiced more often than intervocalic fricatives, and fricatives following a voiceless stop were even more likely to be devoiced. Docherty (1992) for British English, and Veatch (1989) and Stevens et al. (1992), for American, observed similar patterns. In Haggard's and Veatch's results, fricatives in word-final position were virtually always devoiced, comparable to what was observed for fricatives adjacent to voiceless stops. Stevens et al (1992) classify post-vocalic fricatives with at least 30 ms of voicing as "phonetically voiced"; by this criterion, 86% of the phonologically voiced word-final singleton fricatives in their data were "phonetically voiced," a much higher percentage than was found for voiced fricatives preceding (48%) or following (58%) voiceless fricatives.

If at least some of these devoiced fricatives lack any vocal fold vibration at all, the potential might arise for neutralization of devoiced fricatives with voiceless ones. However, work on stop devoicing has shown that devoiced stops do not generally share all the acoustic properties of voiceless stops (Dinnsen & Charles-Luce 1984, Port & O'Dell 1985, Port & Crawford 1989) Perceptual neutralization, meaning that speakers cannot distinguish between devoiced fricatives and their voiceless counterparts, would be most likely if the devoiced fricatives were produced in exactly the same way as the voiceless ones. In fact, English listeners seem to be very successful at distinguishing /s/ and /z/, whether the /z/ is devoiced or not (Stevens et al 1992). The present study is concerned only with the production, not the perception, of fricatives; however, this investigation of the details of production of devoiced fricatives may highlight how listeners are able to distinguish them from voiceless ones.

2. METHOD

The goal of this experiment was to investigate the devoicing of /z/ in a variety of phonological environments in natural speech. These environments were chosen to allow comparisons involving several of the factors suggested by previous work. In order to achieve the most natural sample of speech, the target fricatives were produced in meaningful sentences. In these sentences, /s/ and /z/ occurred in contexts matched for type of neighboring sounds and position in word or phrase. While it was not always possible to match the contexts for /s/ and /z/ exactly, they were as similar as possible with respect to the phonological factors and usually also syntactic position. The word or words that constitute the immediate context for a target /s/ or /z/ will be referred to as an utterance, and one repetition of an utterance will be called a token. Note that in this usage, an utterance consists of only one or two words and is much smaller than a sentence. The experiment included 20 utterances containing target /z/ and 20 containing target /s/. Thus for each context in which a /z/ was measured, an /s/ in a matching context was measured as well. In order to reduce the duration of the experiment, two or three target fricatives were included in each sentence. However, the utterances for each pair of contexts matched for /s/ and /z/ occurred in different sentences. The complete set of sentences is given in Table I.

Factors that were varied in the experiment can be divided between those affecting the local context of the fricative and those that relate to the phrasal context or prosodic position. Both local and prosodic factors were varied to produce different contexts for the fricatives. Local context includes the identity of the following sound and of the preceding sound (see Table II). Prosodic position was either syllable-final, word-final, or sentence-final. Another prosodic factor that was varied was the presence or absence of stress on the syllable containing the fricative. The experiment also included fricatives produced in additional contexts that did not specifically test the effect of the factors listed here; these fricatives were included in comparisons between /s/ and /z/ but not in the analyses of contextual factors.

Table I. Sentences used in the experiment. The underlining has been added to mark the fricatives that were measured.

Ms. Barnes observed him reading this book while he was eating dessert.
Her husband wears a false beard that slides around when he sits down.
John's boss bemoaned his false pretenses for avoiding work.
The music paused for a long time after these bands finished playing.
Pour that liquid into the red sink, and make sure the zinc closure its tightly.
We should replace broken glass from the earthquake before any more of it falls.
The red zinc platter in the kitchen belongs to my housebound aunt.
His boss asked him why he falls behind in his work so often.
The jack in the box pops out very quickly.
When Bob's out, the noise level falls perceptibly.
The statement "Niagara Falls is in Vermont" is totally false.
On a test question, choosing true or false is easier than multiple choice.
There was a short pause before she answered her boss.
Mary's boss laughed for five minutes without a pause.
The long pause outraged impatient listeners in the Roseland concert hall.
The hushed pause lengthened as the Mafia boss passed down Rossland Avenue.
After the protester shouted obscene slogans at the palace guard, he escaped through an ingenious deceit.
The pitcher's lengthy pause postponed the start of the Dodgers game.
A lunar cycle recurs basically once every 28 days.

Table II. Utterances recorded in the experiment. Listed here are the immediate contexts for the target fricatives, which were spoken as part of the complete sentences listed in Table I. Utterances in the same row share the same value for the factor listed at left. Comparisons between /s/ and /z/ used pairs of utterances that were identical for all factors. In some cases the fricatives in a given pair of utterances were used in more than one comparison.

Local context

	/s/		/z/	
Following sound	*preceding /l/*	*preceding vowel*	*preceding /l/*	*preceding vowel*
voiceless stop	false pretense	boss passed	falls perceptibly	pause preserved
voiced stop	false beard	boss bemoaned	falls behind	pause before
vowel	false is	boss asked	falls is	pause outraged
Preceding sound	*syllable-initial*	*word-initial*	*syllable-initial*	*word-initial*
voiced stop	obscene	red sink	observe	red zinc
vowel	deceit	he sits	dessert	the zinc

Prosodic context

position in utterance	/s/	/z/
syllable-final	housebound	husband
word-final	boss bemoaned	pause before
sentence-final	boss.	pause.
stress pattern		
in unstressed syll	palace guard	Dodgers game
in stressed syll	replace broken	recurs basically

Other pairs of words used for overall comparisons

	/s/	/z/
sentence-final	false.	falls.
word-final	pops out	Bob's out
word-final	boss laughed	pause lengthened
word-final	this book	these bands
syllable-final	Rossland	Roseland
syllable coda	test	paused

2.1 Recording technique

Four speakers were recorded in this experiment. They were young adults (20's and 30's) from the Midwest and Western United States. Speakers 1 and 3 were male, Speakers 2 and 4 were female. The sentences were presented to the speakers on individual cards. In cases where an error or pause was detected, speakers were asked to repeat the sentence. Three of the four speakers read the set of sentences six times. For Speaker 2, one token of one utterance was discarded before analysis because of speaker error. For Speaker 3, one token each of three different utterances were discarded. No tokens were discarded for Speaker 4. The remaining speaker, Speaker 1, read the set of sentences five times; however, due to recording problems, 14 sentences were unusable in his data. These sentences included one repetition each of 16 different utterances with target /z/; one additional token containing /z/ was discarded due to speaker error. Additional tokens of utterances with /s/ were also missing from the recording, and one other utterance with /s/ was excluded due to speaker error.

Each speaker wore a pneumotachographic mask to measure airflow and an electroglottograph (EGG) to measure vocal fold contact. These signals and the acoustic signal from a head-mounted microphone were recorded directly to disk at an 8 kHz sampling rate. The airflow and EGG signals were low-pass filtered at 1 kHz, the acoustic signal at 3 kHz. A tape recording was also made. It was digitized at 20 kHz and filtered at 8 kHz for acoustic analysis. Despite the presence of the mask, the frequency information in the acoustics was good to at least 7 kHz.

Two measurements were made on the airflow data. One was the maximum value for any sample during the frication. The other was the mean flow averaged over all samples during the frication. The airflow measurements were calibrated in ml/sec using a flow manometer. The duration of the acoustic segments for the target fricative and the preceding vowel were measured from the waveform of the 20 kHz acoustic signal. In cases where the segmentation was not obvious from the waveform, a spectrogram display on a Kay Elemetrics' CSL system was used as an additional aid. The onset of the preceding vowel was defined as the time when formant structure became apparent; the vowel offset was when the formant structure ended. The onset of the fricative was defined as the time when high-frequency noise became salient in the signal and the offset as the end of this noise. The offset of the vowel was, in some tokens, measured at a distinct, earlier time than the onset of the fricative, in cases where there was an interval lacking either obvious vocalic formant structure or frication noise.

In a subset of the utterances, the fricative was immediately preceded by a stop rather than a vowel; in these cases no measurement was made of the duration of a preceding vowel. These utterances were: "obscene" and "observe", "red sink" and "red zinc", "palace guard" and "Dodgers game", "pops out" and "Bob's out". In utterances including the words "false" and "falls", in which the sequence /ɑl/ preceded the fricatives, the total duration of the /ɑl/ sequence was included in vowel duration, since in most cases it was not possible to determine accurately a boundary between the /ɑ/ and /l/. Occurrences of this sequence occurred in matched utterances with /s/ and /z/ so that measures of the vowel (+[l]) durations were comparable.

2.2 Measurement of amount of voicing

In order to facilitate comparisons among different tokens, it was desirable to identify specific times as the offset or onset of voicing. In measuring voicing, various methods are available that detect vocal fold vibration directly (photoglottography, electroglottography) or indirectly (acoustic measures). Electroglottography (Colton & Conture 1990) was chosen for this experiment because it permits accurate detection of vibration of the vocal folds while being entirely non-invasive, which makes it possible to record speech samples more representative of a speaker's normal behavior than a more invasive technique would allow. It also permits accurate detection of low-amplitude vocal fold vibration even in the presence of acoustic noise, and makes it possible to measure the amplitude of vibration.

One problem in measuring vibration of the vocal folds is that the changes in the signal are gradual, which makes it difficult to select a specific moment as the beginning or end of vocal fold vibration (Docherty 1992). To ensure systematicity in the measurements, the following algorithm was used to identify where vocal fold vibration was present in the EGG signal. The amplitude of one EGG cycle (maximum – minimum during one excursion) was measured at the time of maximum acoustic RMS energy in the vowel preceding the fricative, as calculated by the ENERGY command in CSL. In utterances in which a stop preceded the fricative, maximum RMS energy was measured in the vowel preceding the stop. The fricative was considered to be voiced during the portion of its duration that the amplitude of the EGG cycles exceeded one-tenth of the EGG cycle amplitude at the time of maximum energy in the preceding vowel. Voicing was considered to cease when the amplitude of an EGG cycle fell below this criterion. If the amplitude of several successive EGG cycles wavered between just above and just below the criterion, the offset of voicing was marked where the average amplitude of two successive cycles was below the criterion. For each token of /z/, the percentage of fricative duration with voicing was calculated by dividing the duration of frication with EGG amplitude exceeding the criterion by the total duration of acoustic frication.

The tokens of /z/ were divided into three categories according to the percentage of their duration during which there was voicing. The three categories were:

0 - 25% voicing	devoiced
25 – 90% voicing	partially devoiced
90 - 100% voicing	voiced

Since each token was categorized individually, it was possible for the several tokens of a given utterance to fall into different voicing categories. These category boundaries were chosen because they seemed to reflect the grouping of the data. In particular, the data for Speakers 1, 2 and 3 were bimodal, with few tokens of /z/ having 50-90% voicing, but a number of tokens having over 90% voicing – almost but not quite all of the duration of the frication. It seemed reasonable to group these tokens with over 90% voicing in the "voiced" category together with those that were voiced throughout the entire duration of frication. The boundary between the categories was less clear for Speaker 4. There was not a very clear boundary between devoiced and partially devoiced categories for any speaker, but the 0-25% division grouped together most of the tokens with less voicing.

2.3 Numerical analyses

Each speaker's data were analyzed separately. Two kinds of analyses were performed. The first analysis compared the physiological measures of /s/ and /z/. Each token containing /z/ was paired with the corresponding token containing /s/, that is, the matched utterance produced in the same repetition of the sentences. These pairs of tokens were grouped into the voicing categories described in section 2.2 according to the percentage of voicing in the token with /z/. For example, for Speaker 2 the second repetition of sentence-final "falls" was paired with the second repetition of sentence-final "false". Since there was no vocal fold vibration during the /z/ in this repetition of "falls", these tokens are both tallied in the "devoiced" category. This method ensures that the comparisons between /s/ and /z/ are being made between tokens that occurred in matching contexts. However, the comparisons involve different numbers of tokens, because the different speakers produced different numbers of tokens that were devoiced, partially devoiced or voiced. Within each voicing category, paired t-tests were used to compare the acoustic and aerodynamic measurements of the /s/ tokens with the measurements of the matching /z/ tokens. Therefore, for this analysis only those tokens where both members of the pair were available could be included. However, a few tokens were included in the comparison of acoustic durations where the airflow signal was unusable but the acoustic signal was available. A consequence of this procedure is that in a few cases the number of tokens analyzed was different for the acoustic measurements than for the airflow measurements.

The other analysis investigated the likelihood of devoicing in different contexts. The count of the number of tokens in each of the voicing categories provided the data for this analysis. No statistical tests were done because of the small numbers of tokens. When a token containing /z/ was available, but its matching token with /s/ was not, the token with /z/ was included in this analysis even though it had not been used in the first analysis.

3. RESULTS

3.1 Comparison of acoustic durations and airflow for /s/ and /z/

Paired utterances containing /s/ and /z/ were compared separately for each of the three voicing categories. The measurements that were compared were acoustic duration of the fricative noise and of the vowel preceding the fricative, and for the mean and maximum airflow during the fricative. Mean values and results of the t-tests for the measures of acoustic duration are shown in Table III. Fricative duration was significantly shorter for /z/ than for /s/ with p<.01 for all comparisons except the fully voiced /z/'s of Speaker 4, where the significance level was only p<.05. The shorter durations of acoustic frication for /z/ compared to /s/ are probably a consequence of differences in the interaction of aerodynamic conditions and the position of the vocal folds, since the timing of the supralaryngeal constriction is reported to be similar for /z/ and /s/ (Scully 1971).

Table III. Results of t-tests comparing acoustic durations for utterances with /s/ and utterances with /z/. *n* is the number of pairs compared. *t* is the statistic obtained from a Paired t-test, and *p* is the significance level of this statistic.

Fricative duration

speaker	voicing category	mean (ms) utterances with /z/	mean (ms) utterances with /s/	*n*	*t*	*p*<
1	voiced	70.5	115.7	36	-8.328	*0.001*
	partially devoiced	67.6	87.1	25	-3.460	*0.01*
	devoiced	91.0	130.1	21	-4.979	*0.001*
2	voiced	47.4	89.2	15	-6.145	*0.001*
	partially devoiced	56.1	86.0	28	-6.124	*0.001*
	devoiced	71.3	90.8	76	-6.225	*0.001*
3	voiced	73.9	94.2	15	-3.088	*0.01*
	partially devoiced	57.0	82.9	52	-6.400	*0.001*
	devoiced	79.9	148.1	50	-11.322	*0.001*
4	voiced	64.9	88.9	8	-3.321	*0.05*
	partially devoiced	62.6	85.5	56	-7.636	*0.001*
	devoiced	81.8	113.6	56	-9.707	*0.001*

Duration of preceding vowel						
1	voiced	140.9	121.7	29	3.545	*0.01*
	partially devoiced	171.8	145.7	18	3.967	*0.001*
	devoiced	195.4	150.9	20	3.459	*0.01*
2	voiced	125.7	120.8	9	0.428	*ns*
	partially devoiced	182.6	150.5	23	4.183	*0.001*
	devoiced	216.2	164.0	63	9.034	*0.001*
3	voiced	139.3	133.0	11	1.103	*ns*
	partially devoiced	143.0	129.4	45	2.911	*0.01*
	devoiced	200.0	84.3	33	15.048	*0.001*
4	voiced	127.5	129.1	8	-0.147	*ns*
	partially devoiced	137.7	108.6	45	5.890	*0.001*
	devoiced	178.1	134.7	43	5.289	*0.001*

In English, vowels are typically longer before voiced obstruents than before their voiceless counterparts. The vowel was significantly longer before /z/ than before /s/ (p<.01) for all speakers' devoiced and partially devoiced /z/'s, and for Speaker 1's voiced /z/'s. For Speakers 2 and 3, vowels preceding voiced /z/'s tended to be longer than vowels preceding the matched /s/'s, but the difference was not significant. For Speaker 4, there was an insignificant difference in the duration of the vowels such that they were slightly shorter before voiced /z/'s than matched /s/'s. With this one exception, the vowel durations showed the expected duration difference, whether or not there was vocal fold vibration present in the /z/'s. For Speakers 1, 2 and 3, vowels preceding voiced /z/'s were longer than the vowels preceding the matching /s/'s; this difference was amplified for vowels preceding devoiced /z/'s, where the difference in vowel duration is greater for all speakers. The larger durational differences in the vowels before devoiced /z/'s may support the distinction between /z/ and /s/ in the absence of vocal fold vibration.

The results of the t-tests comparing measurements of mean and maximum flow are shown in Table IV. Again, these results show significant differences between the /z/'s and /s/'s, with /z/'s characterized by generally lower airflow than /s/'s. An important point is that devoicing did not neutralize the distinction between /s/ and /z/. For the mean airflow, used here to characterize the overall level of aerodynamic effort associated with a fricative, there were significantly lower values (p<.01) for /z/ than for corresponding /s/ in 9 out of the 12 comparisons. The exceptions were Speakers 2 and 4's fully voiced /z/'s and Speaker 4's devoiced /z/'s. Speaker 4 was exceptional only in that the comparisons involving her devoiced /z/'s and fully voiced /z/'s did not reach significance; mean airflow was lower for /z/ than for /s/ in both of these comparisons. However, for Speaker 2's fully voiced /z/'s, the mean flow was a non-significant 0.6 ml/sec greater than the corresponding /s/'s, the opposite of what would be expected. Since this comparison, based on only 15 data points, was in the opposite direction from all the other comparisons, the overall pattern of lower mean airflow for /z/ is nonetheless strongly supported, even for Speaker 2. Similarly, maximum flow was significantly lower for /z/ than for /s/ (p<.05) for all voicing categories for all speakers except for the few fully voiced /z/'s produced by Speaker 2. In these, the maximum flow was lower for /z/ than for /s/ but the difference was non-significant.

Table IV. Results of t-tests comparing airflow for utterances with /s/ and utterances with /z/. *n* is the number of pairs compared. *t* is the statistic obtained from a Paired t-test, and *p* is the significance level of this statistic.

Mean airflow (ml/sec)

speaker	voicing category	mean for utterances with /z/	mean for utterances with /s/	*n*	*t*	*p*<
1	voiced	50.3	82.3	32	-5.302	*.001*
	partially devoiced	47.5	83.4	23	-4.277	*.001*
	devoiced	40.0	73.4	21	-3.266	*.01*
2	voiced	100.4	99.8	15	0.057	*ns*
	partially devoiced	89.5	125.0	28	-4.523	*0.001*
	devoiced	96.4	124.4	76	-5.495	*0.001*
3	voiced	71.5	123.8	15	-4.312	*0.001*
	partially devoiced	83.2	137.4	52	-9.168	*0.001*
	devoiced	102.7	128.5	50	-3.636	*0.001*
4	voiced	129.8	142.8	8	-1.674	*ns*
	partially devoiced	141.0	166.1	56	-4.277	*0.001*
	devoiced	150.5	159.7	56	-1.523	*ns*
Maximum airflow (ml/sec)						
1	voiced	138.0	186.1	32	-3.510	*0.01*
	partially devoiced	121.4	198.7	23	-4.556	*0.001*
	devoiced	106.4	201.2	21	-5.571	*0.001*
2	voiced	151.2	162.1	15	-0.674	*ns*
	partially devoiced	131.8	204.5	28	-4.754	*0.001*
	devoiced	138.2	198.4	76	-8.255	*0.001*
3	voiced	148.4	229.6	15	-3.231	*0.01*
	partially devoiced	153.8	254.7	52	-7.700	*0.001*
	devoiced	195.9	238.7	50	-3.453	*0.01*
4	voiced	184.6	226.0	8	-2.387	*0.05*
	partially devoiced	187.8	239.4	56	-5.715	*0.001*
	devoiced	198.1	246.3	56	-5.379	*0.001*

The airflow data thus coincide with the durational results in showing that /z/ and /s/ are different, whether or not the /z/'s are produced with vocal fold vibration. The lower airflow commonly observed for (voiced) /z/ than for /s/ may be due to the vocal folds' more approximated position during /z/ (Isshiki & Ringel 1964, Scully 1971). This suggests that airflow will be lower for /z/ regardless of the amount of vocal fold vibration. If the glottis is in an intermediate state between its approximated position for consonants with voicing and its open position during voicelessness, then the resistance it offers to transglottal airflow should be greater than during voicelessness.

Another explanation for why airflow is less in devoiced /z/ than voiceless /s/ is that speakers may be controlling airflow independently of other articulatory settings, so that they are producing the appropriate airflow for /z/ even when the vocal fold vibration is absent.

The airflow measurements reported here show that /z/ almost always had lower airflow than /s/, with the only real exception being Speaker 2's fully voiced /z/'s. For the airflow comparisons reported here, the fricatives were grouped by amount of voicing, ignoring the different contexts in which they were produced. However, contextual factors such as syllable position may play a role in determining the amplitude of flow. Many of the utterances tested in this experiment placed /z/ at the end of a syllable, and in particular, 11 of Speaker 2's 15 fully voiced /z/'s were syllable- (or word-) final. Hardcastle & Clark (1981) found consistently lower airflow for /z/ than for /s/ for syllable-initial fricatives, but not always for syllable-final fricatives. Isshiki & Ringel (1964) found more variable airflow in syllable-final position than syllable-initial for their sample of consonants. These findings suggest that the higher than expected airflow for Speaker 2's fully voiced /z/'s may be due to their position in the syllable.

3.2 Likelihood of devoicing in different segmental contexts

Another analysis was carried out to determine what factors determine the likelihood of devoicing in different segmental contexts. The contextual factor that has the largest effect on the likelihood of devoicing is the type of sound that follows the fricative. The effect of following context can be seen by comparing the frequency of devoicing in /z/ across the set of utterances involving the word "pause", followed by a vowel, a sonorant consonant /l/, a voiced stop /b/ or a voiceless stop /p/. These results are shown in Figure 1. The only context in which speakers produced fully voiced tokens was when a sonorant (vowel or consonant) followed the /z/. Complete devoicing of /z/ occurred more often when the following sound was the voiceless stop /p/ than when any of the voiced sounds followed. The likelihood of devoicing tended to increase when less sonorous sounds followed the /z/. Speakers also produced more devoiced tokens of /z/ before a voiceless consonant in another set of utterances with /z/ at the end of "falls", shown in Figure 2. Increased frequency of devoicing in fricatives preceding voiceless sounds has been previously reported by Docherty (1992), Stevens et al (1992), Veatch (1989), and Haggard (1978).

In order to compare speakers' likelihood of devoicing in different contexts, a 'devoicing index' was calculated for the set of utterances that included the word "pause" with different following contexts. This index was designed to provide a single number that indicated how often a speaker voiced the /z/ in a particular utterance. A value of this index was calculated for each speaker for each utterance. The percentage of tokens with full voicing was calculated by dividing the number of tokens of an utterance that a speaker produced with full voicing by the total number of tokens that the speaker produced of that utterance. The percentage with partial devoicing and the percentage with complete devoicing were calculated in the same way. The speaker's devoicing index for that utterance was equal to one-half the sum of the percentage with full voicing times two, plus the percentage with partial devoicing. Completely devoiced productions were given a value of 0. Thus the index value for a speaker who fully voiced every token of an utterance would be 1, for a speaker who partially devoiced every token would be 0.5, and for a speaker who completely devoiced every token would be 0. For example, for the /z/ in "pause outrage", Speaker 3 produced 1 voiced token, 2 partially voiced, and 2 devoiced. His index value for this utterance equals $^{1}/_{2} * ((2 * ^{1}/_{5}) + (1 * ^{2}/_{5}) + (0 * ^{2}/_{5})) = 0.4$.

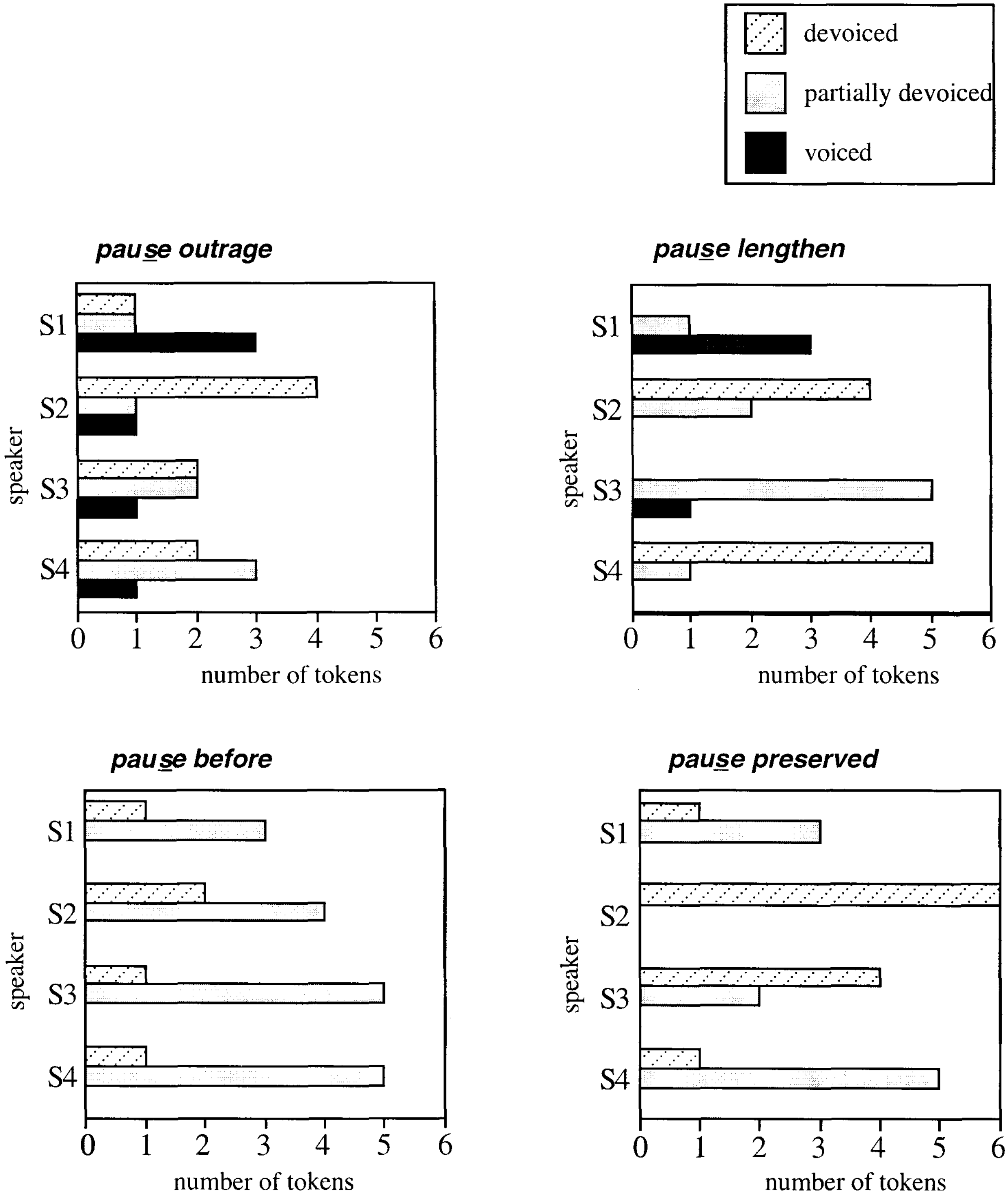
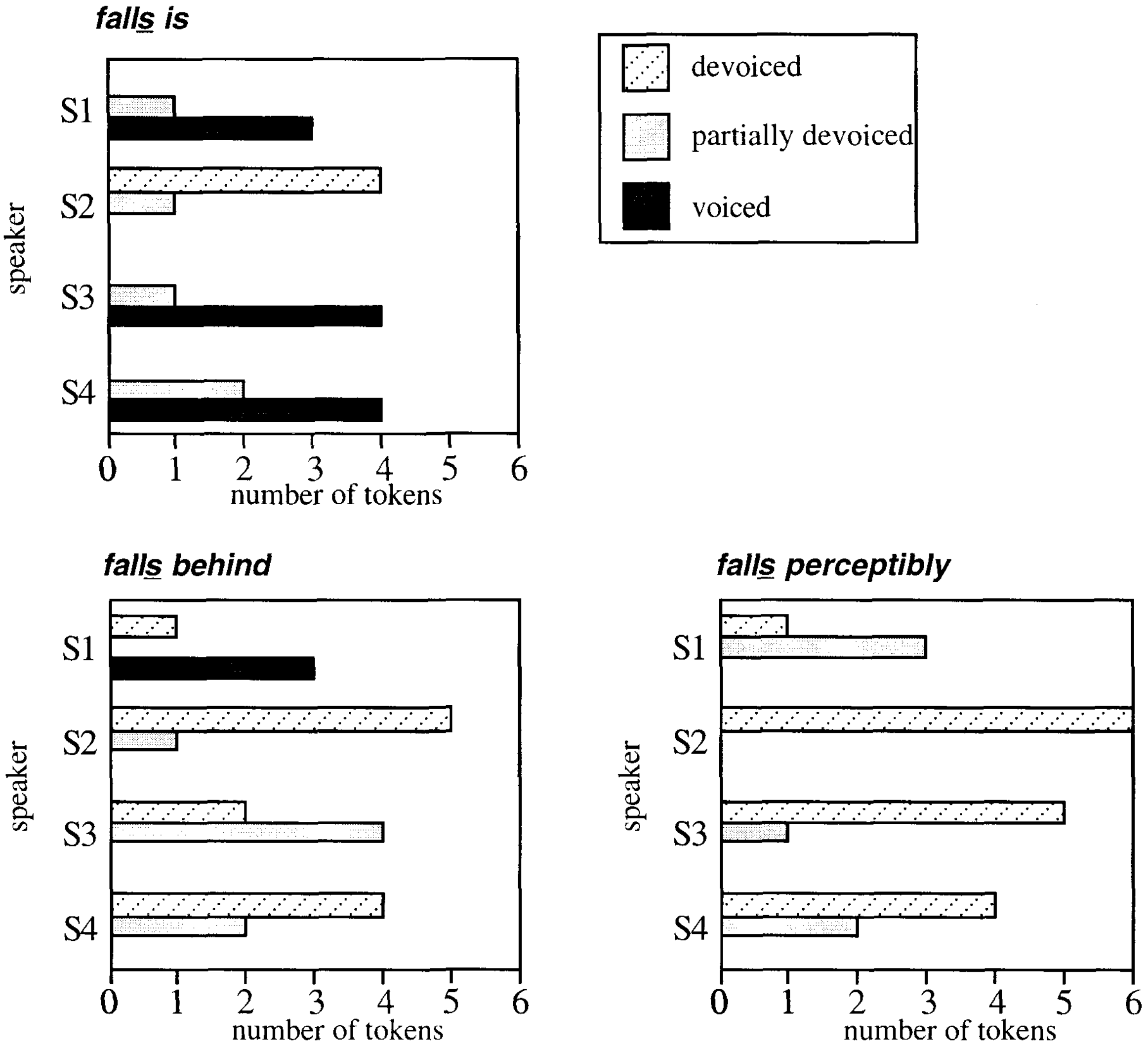


Figure 1. The number of tokens of /z/ in the word "pau*s*e" followed by a vowel, a sonorant consonant /l/, a voiced stop /b/ and a voiceless stop /p/ that the speakers produced as devoiced, partially devoiced, or fully voiced.



falls is

falls behind

falls perceptibly

Figure 2. The number of tokens of /z/ in the word “falls” followed by a vowel, a voiced stop /b/ and a voiceless stop /p/ that the speakers produced as devoiced, partially devoiced, or fully voiced.

The different speakers’ values of this index were rank ordered for each of the utterances that included “pause”, and these orderings were compared to test the prediction that decreased sonority of the context following the /z/ made devoicing more likely. The expected rank ordering was that the most devoicing would occur before /p/, with less before /b/, even less before /l/, and the least of all before a vowel. The number of speakers who conformed to these predictions are listed below.

- all speakers had devoicing more often before /p/ than before /b/
- all but Speaker 4 had devoicing more often before /p/ than before /l/
- all speakers had devoicing more often before /p/ than before a vowel
- Speakers 1 and 3 had devoicing more often before /b/ than before /l/
- Speakers 1 and 4 had devoicing more often before /b/ than before a vowel
- Speakers 2 and 4 had devoicing more often before /l/ than before a vowel

No speakers produced any fully voiced tokens of /z/ in "pause" when there was a /b/ following. This may reflect an absence of vocal fold vibration during /b/, which is often produced as a voiceless unaspirated stop by speakers of American English (Lisker & Abramson 1964). No measurements were made in the current study of the voicing of /b/, but inspection of the data suggests that few tokens of /b/ had any vocal fold vibration. Although /b/ probably often lacked voicing, there were nonetheless differences in the likelihood of devoicing between /z/'s preceding /b/ and /z/'s preceding /p/. Devoicing was more likely to occur preceding /p/, which normally has glottal opening rather than adduction as is typical for /b/ (Lisker, Abramson, Cooper & Schvey 1969; Flege 1982).

The context preceding the /z/ also affects the likelihood of devoicing. For comparisons of preceding context, the /z/ was either in a monosyllabic word or at the beginning of a syllable with primary stress. Figure 3 shows that speakers produced more devoiced /z/ when /z/ was preceded by a voiced stop than by a vowel. This pattern held for both syllable-initial and word-initial /z/, but in general there was more devoicing in syllable-initial, word-medial position than in word-initial position. Like the effect of the sound following the /z/, the effect of the sound preceding /z/ appears to be an assimilatory process. This voicing assimilation seems to be operating in both directions. The likelihood of devoicing increased if there was a voiceless context on either side of the /z/; similarly, in fricative clusters in English, a phonologically voiced fricative was more likely to devoice when adjacent to a voiceless fricative than vice versa (Stevens et al 1992). The effect of phonologically voiced stops that are adjacent to /z/ seems to be intermediate between the effects of vowels and of /p/, which is appropriate if there is neither vocal fold vibration as in vowels nor a glottal opening gesture as in /p/.

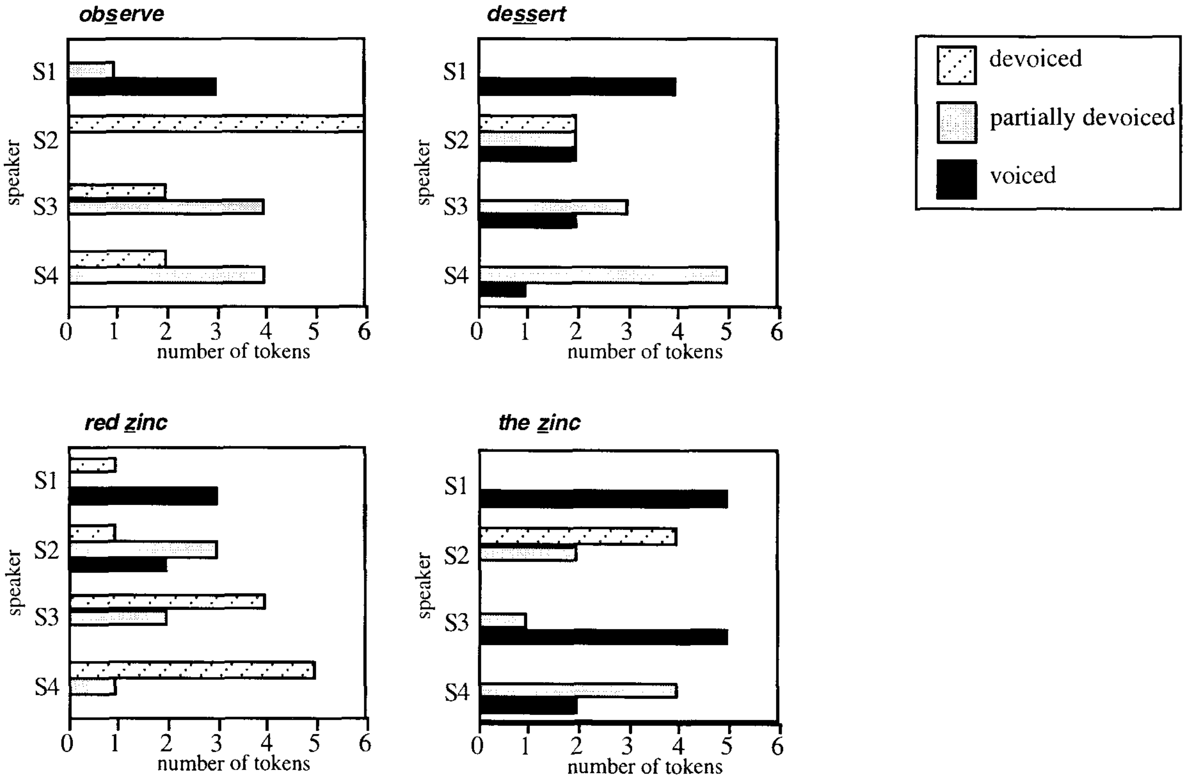


Figure 3. The number of tokens of syllable-and word-initial /z/ preceded by a voiced stop or a vowel that the speakers produced as devoiced, partially devoiced, or fully voiced.

In addition to there being differences in the likelihood of devoicing in different contexts, there were also substantial differences among the four speakers in their overall likelihood of devoicing. Figure 4 shows the number of tokens that each speaker produced with different amounts of voicing. The tokens are grouped according to the voicing categories described in section 2.2. Speaker 1 often produced /z/ with vibration throughout, whereas Speaker 2 produced the most tokens with no vocal fold vibration. Speakers 3 and 4 tended to produce /z/ with vocal fold vibration during part of the fricative. For Speakers 1 and 2, and to some extent Speaker 3, few tokens were produced with 50-90% of voicing, but a larger number had 90-100% voicing. Speaker 4 showed a slightly different pattern: the number of tokens decreased gradually with larger percentages of voicing so that few tokens were produced with very much voicing. These overall differences among the speakers suggest that the individuals may have different global settings for likelihood of devoicing, which also contributed to the differences in likelihood of devoicing in the different contexts that were tested.

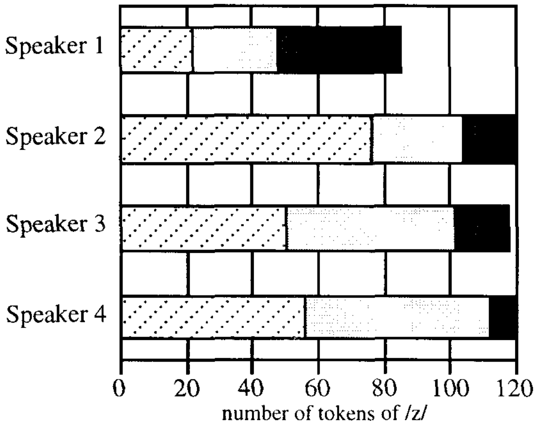


Figure 4. The total number of tokens of /z/ that each speaker produced as devoiced, partially devoiced, or fully voiced.

3.3 Likelihood of devoicing in different stress and phrasal contexts

In addition to the local assimilatory effects described above, influences from prosodic structure were also investigated. One set of words compared /z/'s in sentence-final, word-final and syllable-final position, with a voiced stop following the word- and syllable-final /z/ ("pau<u>s</u>e.", "pau<u>s</u>e before", and "hu<u>s</u>band"). It was expected that the /z/ would be devoiced more often at the end of a larger prosodic unit. This prediction was based on evidence of articulatory reduction at the end of prosodic units, a pattern that has been observed as shrinkage of articulatory movements (jaw opening, in Vayra & Fowler 1992; velum raising in Krakow, Bell-Berti & Wang 1995), reduced tongue-palate contact during a stop in coda position compared to onset position (Keating 1995), and reduced gestural magnitude for the glottis in non-phrase-initial position (Pierrehumbert & Talkin 1992). Following this line of research, devoicing could be considered a form of lenition resulting from lower articulatory or aerodynamic activity at the end of a prosodic constituent. If there were reduction in the oral articulation, giving a more open lingual constriction than usual for /z/, a loss of frication would result. This was not observed in the data, which suggests that reduction is instead taking the form of lower pulmonic effort, less effort at expanding the oral cavity, or relaxation of vocal fold tension. Lenition of consonants is most often considered to involve a change from voiceless to voiced (Lass 1984), especially intervocalically as assimilation to the voicing of neighboring vowels. However, at the end of a sentence, lenition may take the form of assimilation to the voiceless state of a following pause.

The prediction for greater likelihood of devoicing at the end of larger domains was borne out in the data. All sentence-final /z/'s produced by all speakers were completely devoiced. In addition, speakers were more likely to devoice the /z/ word-finally than at the end of a word-medial syllable; however, there was considerable individual variation for the word-medial syllable which can be seen in Figure 5. All speakers produced most of their tokens of word-final /z/ in "pause before" with partial devoicing; no speaker produced any fully voiced tokens of /z/ in this utterance. In contrast, for word-medial /z/ in "husband", Speakers 1, 2 and 3 produced one or more fully voiced tokens; Speakers 1 and 3 also produced some partly or completely devoiced tokens. However, Speaker 4 produced more tokens of devoiced /z/ in word-medial position than in word-final position. Given that all of this speaker's word-medial /z/'s in "husband" were completely devoiced, the most likely interpretation is that she has adopted a pronunciation of "husband" with an /s/ rather than a /z/. Further evidence for this interpretation is that the mean airflow for the /z/ in "husband" was higher than the airflow for the matching /s/ in "housebound." This result is discussed further in section 3.4.

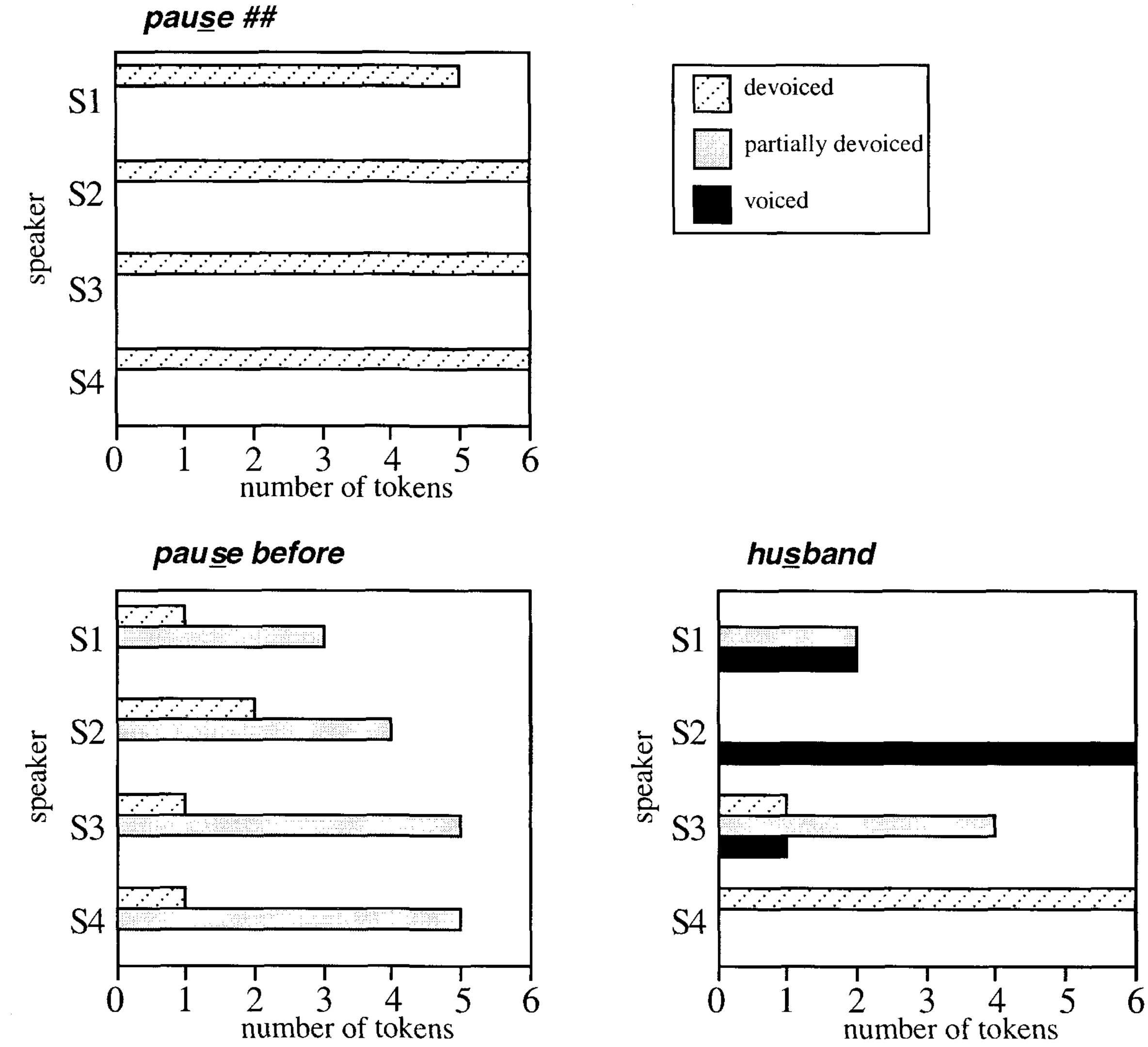


Figure 5. The number of tokens of sentence-, word- and syllable-final /z/ that the speakers produced as devoiced, partially devoiced, or fully voiced.

The effect of stress on the likelihood of devoicing can also be considered a reflection of prosodic organization. Two-syllable words with different stress patterns were compared to see whether the word-final /z/ was more likely to be devoiced at the end of stressed or unstressed syllables. It was expected that devoicing would be more frequent at the end of an unstressed syllable than at the end of an stressed syllable, since the lack of stress signals a prosodically weaker position. Speakers 1 and 3 fulfilled this prediction, as can be seen in Figure 6. Speaker 1 produced the stressed /z/ in "recúr<u>s</u>" either fully voiced or partially devoiced, and the unstressed /z/ in "Dódger<u>s</u>" as either partly or completely devoiced. Speaker 3 produced most tokens of stressed /z/ with partial devoicing, but always produced the unstressed /z/ as fully devoiced (6 tokens). Unlike these two speakers, Speaker 2 made no difference in the likelihood of voicing stressed and unstressed /z/, so her results neither support nor contradict the hypothesis. Only Speaker 4 showed a pattern somewhat contrary to the prediction: she produced equal numbers of partly and fully devoiced tokens of stressed /z/, but for unstressed /z/ she produced more partly devoiced tokens. Averaging across the 4 speakers there were 23% fewer tokens with partial or full voicing in the unstressed /z/ than in the stressed /z/, supporting the hypothesis that devoicing is more likely in the weaker unstressed syllable.

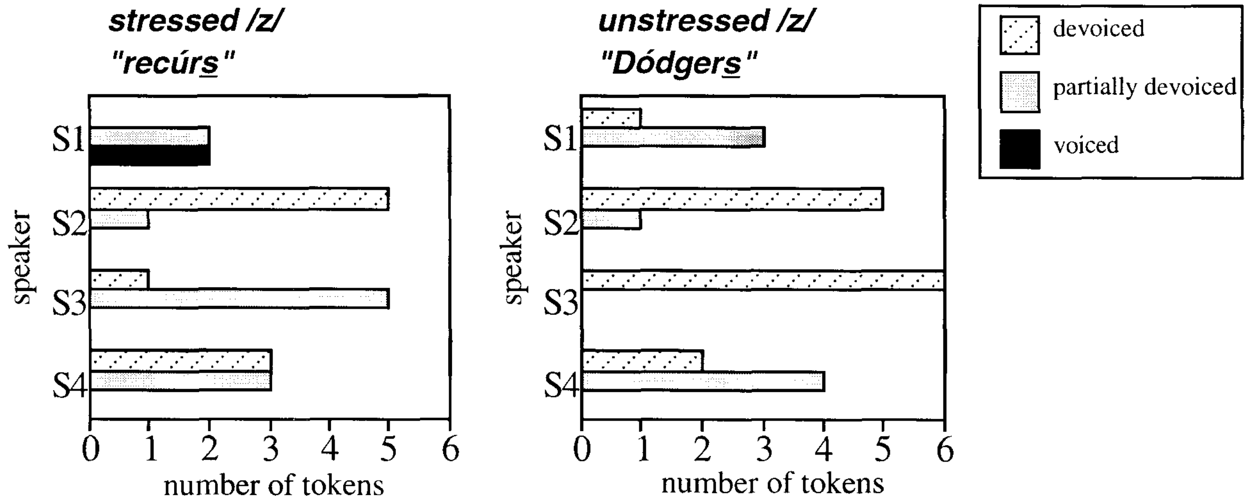


Figure 6. The number of tokens of /z/ at the end of a stressed or unstressed syllable that the speakers produced as devoiced, partially devoiced, or fully voiced.

3.4 Comparison of airflow in different utterances

The t-tests reported in section 3.1 showed that for voiced, partially devoiced and completely devoiced /z/'s, airflow was lower for the /z/'s than for the comparable /s/'s. In these comparisons, tokens were grouped by percentage duration of voicing, combining tokens of different utterances within each voicing category. Other comparisons were made grouping tokens of the same utterance regardless of duration of voicing, and comparing the airflow for tokens of different utterances. One set of comparisons looked at the difference between /p/ and /b/ contexts following /z/. If devoicing of /z/ preceding /p/ is the result of (glottal) assimilation during /z/ to the open glottis of /p/, higher airflow might be expected in /z/'s that are completely devoiced before /p/ than in those that were not completely devoiced. Devoicing of /z/ before /b/ is hypothesized to be the consequence of insufficient subglottal pressure; lower airflow might be found in these devoiced /z/'s if the reduced pressure is due to less air being expelled from the lungs. The data show that before /b/, there was lower maximum and mean airflow in /z/'s that were completely devoiced than in /z/'s that were fully voiced or partially devoiced. This suggests that there is lower pulmonic pressure in the devoiced /z/'s, which may be responsible for the devoicing. Before /p/, there were no tokens of fully voiced /z/, and there was much variability as to whether completely or partially devoiced /z/ had greater flow. This variability suggests that the devoicing in this case could have resulted from either increased glottal opening (with greater flow) or from insufficient subglottal pressure (with less flow).

To investigate further whether devoicing can be attributed to lower aerodynamic effort, comparisons were made of the airflow values for the utterances in different prosodic contexts grouped by phrasal context (syllable-final, word-final or sentence-final) rather than by voicing category. The patterns for maximum flow and mean flow were similar; the data for maximum flow are given in Figure 7. For Speaker 1, the flow is lower for /z/ than for /s/ in each of the three contexts. Since all the sentence-final /z/'s were completely devoiced, but some of the syllable-final ones were fully voiced, it can be concluded that for Speaker 1, regardless of the amount of voicing in the /z/, the airflow for /z/'s is lower than that of /s/'s produced in the same context. For this speaker, there is no neutralization between /z/ and /s/.

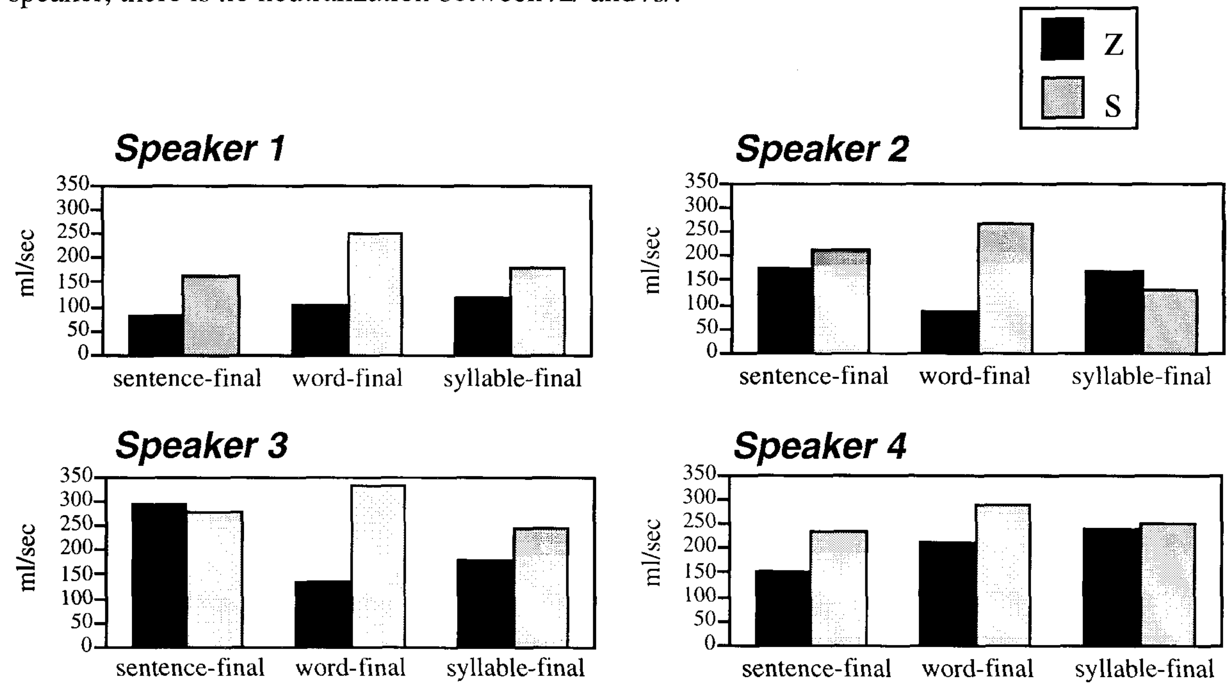


Figure 7. Maximum airflow during /z/ and /s/ in different prosodic positions: sentence-final "pau<u>s</u>e", word-final "pau<u>s</u>e before" and syllable-final "hu<u>s</u>band".

On the other hand, Speaker 3 showed lower airflow for /z/ than for /s/ in word- or syllable-final position, as expected, but surprisingly shows relatively high flow for both /z/ and /s/ in sentence-final position. This high air flow suggests that the speaker may be neutralizing the /z/-/s/ contrast in this one position. It may be that this speaker is opening the vocal folds at the end of the sentence in anticipation of the open position of the glottis that is typical of respiration during a pause. The high airflow in the sentence-final position appears to be a special case of an assimilatory process, different from the examples of assimilation to adjacent sounds that were discussed earlier.

The pattern in the flow data is almost the opposite for Speaker 4 than for Speaker 3. For Speaker 4, airflow is lower, as expected, in the sentence-final and word-final /z/'s than in the corresponding /s/'s, but the airflow in the syllable-final /z/'s is almost as high as in the syllable-final /s/'s. Speaker 4 always completely devoiced these syllable-final /z/'s in the word 'hus̲band'. The high airflow for these devoiced /z/'s suggest that, as noted earlier, this speaker has a different lexical form for this word, so that it contains an /s/ rather than the /z/ that might be expected. Such lexical variation is found in English in other words, such as the two pronunciations 'ab[s]urd' and 'ab[z]urd'. The airflow data for Speaker 2 was similar to that of Speaker 4, but the high airflow for syllable-final /z/ is difficult to explain for Speaker 2 since her /z/'s were fully voiced in this utterance.

4. DISCUSSION

4.1 Two mechanisms for devoicing

Two separate influences on the likelihood of devoicing of /z/ have been identified here: one that can be viewed as assimilation to an adjacent context that lacks voicing, and one that can be ascribed to the position of the /z/ in phrasal or prosodic structure. The results demonstrate that while much of the variability in the likelihood of devoicing can be accounted for by these influences combined with speakers' overall differences for likelihood of devoicing, devoicing in these /z/'s is nonetheless a process best described in probabilistic terms – more or less likely rather than possible or impossible.

As discussed in the introduction, voiced fricatives present a particularly exigent set of demands on the vocal tract. Because they require precise conditions for successful production, it may be that a comparatively small divergence from these conditions is more likely to result in a salient difference from the "default" characteristics of a voiced fricative than would be the case for some other sound. In the present data, divergence from the canonical form of /z/ always showed up as devoicing, rather than loss of frication. The tendency to devoice can be explained in part by evidence suggesting that the glottis is always somewhat open during voiced fricatives – more open, at least, than for voiced stops. Such evidence comes from studies using transillumination to examine glottal opening (Lisker, Abramson, Cooper & Schvey 1969), as well as EMG data showing more suppression of the adductory interarytenoid and lateral cricoarytenoid muscles in word-medial voiced fricatives than voiced stops (Hirose & Ushijima 1978).

If the usual state of the glottis for voiced fricatives is somewhat open, maintaining sufficient subglottal pressure for vocal fold vibration will require greater airflow from the lungs than for voiced sounds produced with a more closed glottis, so vocal fold vibration may fail more often. In addition, just a small additional opening of the glottis could lead to devoicing. Laver (1994) argues that during devoiced sounds such as /z/ the glottis is probably in a state intermediate between voicing and voicelessness, like the state of the glottis that is used in whisper, with the glottis open but the folds very close together. The glottis also appears to be somewhat open for devoiced vowels in Japanese, unlike voiced productions of the same vowels (Hirose 1971). There are two reasons to suppose that the glottis may behave differently in the Japanese devoiced vowels than in the voiced fricatives examined here. Voicing is not contrastive for vowels in Japanese, so there is no risk of neutralization between two contrastive sounds of the language if vowels are produced with an open glottis, whereas in English there is a contrast between phonologically voiced and voiceless fricatives, so that it may be perceptually advantageous for devoiced fricatives to remain distinct from voiceless ones. It is also much easier to maintain vocal fold vibration during a vowel than during a (voiced) fricative; the glottis may have to open slightly in order to stop vocal fold vibration during a vowel, whereas during a fricative vocal fold vibration seems to stop very easily. Thus, the Japanese evidence does not argue conclusively that devoicing in English fricatives results from active glottal opening.

Based on the results of the experiment reported here, there appear to be two possible routes to devoicing: in one scenario the glottis opens more, increasing trans-glottal airflow and equalizing pressure above and below the glottis so that the vocal folds no longer vibrate. Alternatively, the configuration of the glottis remains the same, but sub-glottal pressure nonetheless becomes insufficient to maintain vibration of the folds. The second scenario is plausible in part because muscular action is necessary to maintain adequate subglottal pressure for speech (Ohala 1990). Thus if a speaker does not use enough effort in the abdominal muscles, subglottal pressure will fall below the level necessary to keep the vocal folds vibrating. This second scenario relies, therefore, on a reduction of effort by the speaker, whereas the first scenario (favored by Haggard 1978), relies on extra effort leading to additional opening of the glottis.

The second scenario seems appropriate to explain cases where the voicing for /z/ is lost because of reduced effort, such as seems typical of the prosodic positions where devoicing was found to be most common. The /z/'s in utterance-final or word-final position, or unstressed syllable codas, are in positions where articulator movement is often reduced (e.g. Manuel 1991; Krakow 1993, Byrd 1994, Keating 1995, Browman & Goldstein 1995). However, these data do not distinguish whether this reduction occurs because the final element in a prosodic constituent is maximally affected by declination, or because it does not undergo any initial strengthening (see Fougeron & Keating 1996 for discussion).

In contrast, the /z/'s that are devoiced because they are adjacent to a voiceless sound or pause may require a different explanation. For these, the first scenario, which proposes possible increased glottal opening as the mechanism for devoicing, seems more appropriate. The glottis "assimilates" to the adjoining sound's requirement for an open glottis. These /z/'s seem to be more malleable than the neighboring sounds to which they assimilate. As with the /z/'s whose devoicing is ascribed to their weak position in prosodic structure, this malleability may manifest itself because a small change to the articulatory position or aerodynamic conditions of a /z/ is more likely to have a discernible effect than a correspondingly small change to some other more robust articulation. Even if the "assimilated" /z/'s are devoiced because the glottis is more open, the additional opening cannot be very great because the airflow for these /z/'s is still low compared to /s/.

Opening the glottis in anticipation of a following voiceless sound can also be viewed as a simplification of the articulatory demands of the utterance. In the vast majority of tokens in this experiment (279 of a total 322 devoiced /z/'s), the devoicing was at the end of the /z/; that is, at the beginning of frication the vocal folds were vibrating, but the vibration ceased before the end of frication. If the glottis is indeed opening somewhat during the /z/, this opening is not a separate additional gesture, merely an anticipation of the opening gesture that would otherwise happen at the end of the fricative. This kind of anticipation can be interpreted as a change in the coordination of the glottal and oral gestures: the glottis is opening before the tongue releases its constriction for /z/, rather than simultaneously with the constriction release. Changing the relative timing of two gestures is a common process in "casual speech" as discussed by Browman & Goldstein (1990): no additional gesture is necessary to arrive at a devoiced /z/, just a change in timing of the two gestures involved.

4.2 Devoicing at different times during the fricative

With the exception of Laver (1994), few previous discussions of fricative devoicing have considered the possibility of devoicing at the beginning or in the middle of the fricative rather than at the end. However, all four speakers in the present experiment produced at least a few tokens of /z/ with devoiced portions at the beginning or in the middle. Contrary to the suggestion of Laver (1994), there was no discernible trend for word-initial fricatives to have devoicing at the beginning and word-final fricatives at the end. Devoicing at the beginning was observed in /z/'s in word-initial, syllable-initial word-medial, and word-final positions. Devoicing at the end occurred in the full range of positions tested in the experiment.

Although devoicing at the end of /z/ was by far the most frequent pattern, in 43 tokens the devoiced portion of the /z/ was either at the beginning or in the middle of the acoustic frication. Given that assimilation to the voicing of adjacent sounds is an important determiner for presence or absence of voicing in /z/, devoicing at the beginning of the /z/ would be expected when the preceding sound is voiceless; assimilation would also suggest that the vocal folds are more likely to be vibrating at the end of the /z/ if the following sound is voiced. These predictions are largely correct, assuming that voiced stops in English often lack voicing during the closure (Lisker & Abramson 1964). There were 18 tokens of /z/ with devoicing at the beginning. These included four different utterances, all but one of which had a voiced stop preceding the /z/ and a vowel following. Speaker 3 accounted for 14 of these 18 tokens, Speakers 1 and 4 for two each. Speaker 3's propensity for this pattern of devoicing may be due to a more consistent lack of vocal fold vibration during voiced stops. (Some speakers of English use both prevoicing and short lag VOT in voiced stops, other speakers regularly choose one or the other VOT pattern (Docherty 1992); possibly Speaker 3 uses a short lag more consistently than the other speakers.) Devoicing at the beginning of the fricative may simply be a consequence of an individual speaker's choice of VOT for voiced stops.

Devoicing in the middle of the fricative was produced by all four speakers, and in initial and final syllable and word positions. Speaker 1 produced 2 of these tokens, Speaker 2 produced 7 , and Speakers 3 and 4 produced 8 tokens each. There were also a total of 4 tokens in which voicing ceased briefly in the middle of the /z/ for less than 10% of frication duration so that the token was categorized as fully voiced. The utterances in which speakers devoiced the middle of the /z/ are listed in Table V. These /z/'s occurred with a variety of preceding contexts, and included the four utterances in which devoicing was found at the beginning of the /z/. The medially devoiced /z/'s were all followed by either vowels or in one case [l]. Devoicing in the middle of a /z/ (with voiced frication before and after the devoiced portion) seems incompatible with the proposed assimilatory mechanism, since there is no adjacent glottal opening gesture to instigate assimilation. The devoicing in these tokens must therefore have a different origin. 'Passive' devoicing, such as is hypothesized in prosodically weak environments, could easily occur in the middle of a fricative if a speaker reduced but then re-initiated the muscular effort necessary to expel air from the lungs and create subglottal pressure sufficient for vocal fold vibration. Since greater subglottal pressure is necessary to (re-)initiate voicing after it has stopped than simply to maintain it (Rothenberg 1981), some considerable effort would be required. Another explanation for the loss and re-initiation of voicing is that it may be a consequence of the movement of the tongue forming and releasing the oral constriction. During the fricative, the oral constriction becomes so narrow that intraoral pressure increases enough to stop the vocal folds from vibrating. As the tongue releases the oral constriction, intraoral pressure decreases and voicing begins anew. This explanation of the medial devoicing seems the most probable as it is not predicated on the speaker overcoming hysteresis of the vocal folds. Thus it appears that a change in pressure, either subglottal or oral, is likely to be the mechanism involved when there is devoicing in the middle of a /z/. However, the present data are not sufficient to resolve fully the issue of which devoicing mechanism is occurring in which utterance, especially as the mechanism may not always be the same even for a given utterance.

Table V. Utterances in which one or more tokens were devoiced in the middle of /z/. The total number of tokens of this utterance that were partially or completely devoiced is given in the rightmost column. These values are the totals for all four speakers.

	devoiced in middle	total devoiced
syllable-initial		
observe	2	19
dessert	4	12
word-initial		
red zinc	3	17
the zinc	4	11
syllable-final		
Roseland	2	20
word-final		
Bob's out	5	16
falls is	3	8
pause outrage	2	16

5. CONCLUSIONS

In American English, a speaker may devoice /z/ in almost any environment, but the likelihood of devoicing varies greatly and depends on the preceding and following segmental context, as well as the position of the /z/ in the prosodic structure of the utterance. Individual speakers also varied as to how often they devoiced; they may be considered to have different settings of a global "likelihood of devoicing" parameter. Each speaker's parameter setting effects an overall shift on the values for likelihood of devoicing that are otherwise determined by local and prosodic context. Across contexts, the rank ordering of individual speakers for likelihood of devoicing was fairly similar. That is, certain contexts were more favorable for devoicing than others for most or all speakers. Further investigation will be necessary in order to identify other influences on the likelihood of devoicing, and whether any of the effects identified here are lexically idiosyncratic.

Comparisons of acoustic durations and airflow measurements showed that even completely devoiced /z/'s did not become identical with /s/'s produced in similar contexts. The /z/'s, whether fully voiced, partially devoiced or completely devoiced, are characterized by lower airflow than /s/, which suggests that the glottis is less open or that less pulmonary effort is expended. The low airflow suggests that speakers are not making an extra opening of the glottis in order to devoice the /z/; this devoicing process seems to be characterized more by a reduction in effort than an addition of another gesture.

The results presented here show much variability in likelihood of devoicing, but a large part can be attributed to differences in context. Tolerance for such variability may be related to the constraints on the speaker, as well as the interaction between speaker and listener (Lindblom 1990). The pattern of reduction in weak prosodic environments suggests that in these environments there is less need for the speaker to produce maximal distinctions between /z/ and /s/; given the exacting requirements for a voiced /z/, a speaker may opt to reduce effort and produce a devoiced /z/ that is more economical of effort, since this can be achieved while still keeping a distinction with /s/ and thus satisfying the perceptual needs of the listener.

In voiced sounds, the output of the vocal tract is a product of the interaction between aerodynamic conditions (subglottal pressure, volume of air expelled from the lungs) and the positions of the vocal folds and the articulator forming a supralaryngeal constriction (Bickley & Stevens 1991). In devoiced sounds, either or both the aerodynamics and articulation are different, so that the output of the vocal tract is altered. It has been proposed here that devoicing in /z/ may result from either an assimilatory process, in which the position of the vocal folds is more open (a change in articulation) or from a lenition process, in which the subglottal pressure or volume of air is insufficient to maintain vocal fold vibration (a change in aerodynamics). A third possibility is that the articulatory movement for the supralaryngeal constriction causes a change in the aerodynamic conditions, resulting in devoicing. The present data do not determine which of these changes may occur together, but clearly there are limits to the extent that speakers diverge from an 'ideal' voiced /z/. Speakers of English seem to diverge only by losing voicing, not by losing frication. In addition, the airflow results show that speakers do not diverge so much that the devoiced /z/ acquires all the characteristics of /s/.

The devoicing process examined here is a complex example of the kind of constrained variability that is typical of speech production. Some examples of variability in production that result from "casual speech" processes have been successfully modeled in Articulatory Phonology as changes in the magnitude or timing of articulatory gestures (Browman & Goldstein 1986, 1990, 1992). Devoicing as assimilation would be easy to model in this framework as a change in temporal coordination among gestures on different tiers. Assimilation of voicelessness could be modeled as a glottal opening gesture timed to begin earlier at the formation of a (stop or) fricative constriction.

For the lenition form of devoicing described in this paper, the likelihood of devoicing was found to vary according to whether the gesture is at the end of a word or phrase, i.e. a larger or smaller prosodic constituent. Modeling the lenition type of devoicing requires modulation of articulation and/or aerodynamics over several gestures, unlike contextual effects previously modeled in Articulatory Phonology which typically involved neighboring gestures (Browman & Goldstein 1990, 1992). It may be possible to model some of the lenition effects by changing the phasing of domain-initial and domain-final gestures to reflect the differences in their magnitude (Browman & Goldstein 1995). Distinguishing increasing degrees of lenition in gestures that end increasingly large prosodic domains would require cumulatively greater phasing differences.

Alternatively, effects of prosodically-structured variation could be modeled in Articulatory Phonology by a "prosodic" tract variable. One "gesture" for the prosodic tract variable would be associated with each level in the prosodic hierarchy (Nespor & Vogel 1986) that is relevant in a particular utterance. The parameters of such a tract variable would be a function of the number of gestures in the corresponding prosodic constituent (such as word or phrase) and the rank of this constituent in the prosodic hierarchy. The effect of these tract variables would be cumulative, so that at the end of a higher-ranked (larger domain) constituent the gestural parameters would also be affected by the prosodic tract variables for lower-ranked constituents (cf. Fougeron & Keating 1996). At the beginning of a constituent this prosodic tract variable would increase gestural magnitudes from the values that would be determined by local influences; towards the end of a constituent it would reduce them (Vayra & Fowler 1992; Krakow, Bell-Berti & Wang 1995).

The time course of a prosodic tract variable cannot be described too precisely, since it is not clear whether larger amplitudes at the beginning are due to strengthening of the beginning or weakening of the end (or both) (see Fougeron & Keating 1996). However, the critically-damped trajectory associated with articulatory tract variables might well be appropriate if Fougeron and Keating (1996) are correct in their view that strengthening of the initial articulation is more important than weakening of later articulations: the prosodic tract variable could start with a high value for articulatory strengthening and diminish over time along a critically-damped path to a low value corresponding to an absence of strengthening. Since a prosodic tract variable would affect the parameter values of both articulatory and aerodynamic tract variables, different settings of these could model different sources of devoicing. Whether or not the vocal folds actually cease to vibrate in a particular production would be a consequence of the interaction of stress and gestural context as well as the effect of the prosodic tract variable on the parameters of the glottal and supralaryngeal gestures.

Despite the variability observed among speakers and in different contexts, devoicing is not a random process. It is a function of different levels of linguistic organization in an utterance, as well as the interplay between articulatory and aerodynamic conditions in the vocal tract. These tentative suggestions for how the effects of prosodic organization in speech can be integrated with the Articulatory Phonology representation of timing may provide a route to modeling temporal patterns in speech more completely.

ACKNOWLEDGEMENTS

This work was supported by a postdoctoral fellowship from NIH grant DC00008 to the UCLA Division of Head & Neck Surgery. Thanks to Bruce Gerratt and Jody Kreiman for guidance with data collection and analysis. I would also like to thank the UCLA Phonetics Laboratory for the use of their facilities, the speakers in the experiment for their willing participation, and Ian Maddieson for helpful comments on an earlier draft of this paper.

REFERENCES

Balise, R. & Diehl, R. (1994) Some distributional facts about fricatives and a perceptual explanation, *Phonetica*, **51**, 99-110.

Baum, S. & Blumstein, S. (1987) Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English, *Journal of the Acoustical Society of America*, **82**, 1073-1077.

Bickley, C. & Stevens, K. (1991) Effects of a vocal tract constriction on the glottal source: data from voiced consonants. In *Laryngeal function in phonation and respiration* (T. Baer, C. Sasaki & K. S. Harris, editors), pp. 239-253. San Diego: Singular Publishing.

Browman, C. & Goldstein, L. (1986) Toward an articulatory phonology, *Phonology Yearbook*, **3**, 219-252.

Browman, C. & Goldstein, L. (1990) Tiers in articulatory phonology, with some implications for casual speech. In *Papers in laboratory phonology I: between the grammar and the physics of speech* (J. Kingston & M. E. Beckman, editors), pp. 341-376. Cambridge: Cambridge University Press.

Browman, C. & Goldstein, L. (1992) Articulatory phonology: an overview, *Phonetica*, **49**, 155-180.

Browman, C. & Goldstein, L. (1995) Gestural syllable position effects in American English. In *Producing speech: contemporary issues: for Katherine Safford Harris* (F. Bell-Berti & L. Raphael, editors), pp. 19-33. Woodbury, NY: AIP Press.

Byrd, D. (1994) *Articulatory timing in English consonant sequences*. Ph.D. dissertation, UCLA. In *UCLA Working Papers in Phonetics*, **86**.

Catford, J.C. (1977) *Fundamental problems in phonetics*. Bloomington: Indiana University Press.

Colton, R. & Conture, E. (1990) Problems and pitfalls of electroglottography, *Journal of Voice*, **4**, 10-24.

Dinnsen, D. & Charles-Luce, J. (1984) Phonological neutralization, phonetic implementation and individual differences, *Journal of Phonetics*, **12**, 49-60.

Docherty, G.J. (1992) *The timing of voicing in British English obstruents. Netherlands Phonetics Archives*, **9**. Berlin: Foris.

Flege, J.E. (1982) Laryngeal timing and phonation onset in utterance-initial English stops, *Journal of Phonetics*, **10**, 177-192.

Fougeron, C. & Keating, P. A. (1996) Articulatory strengthening in prosodic domain-initial position, *UCLA Working Papers in Phonetics*, **92**, 61-95.

Haggard, M. (1978) The devoicing of voiced fricatives, *Journal of Phonetics*, **6**, 95-102.

Hardcastle, W.J. & Clark, J.E. (1981) Articulatory, aerodynamic and acoustic properties of lingual fricatives in English, *Work in Progress, Phonetics Laboratory, University of Reading*, **3**, 51-79.

Hirose, H. (1971) The activity of the adductor laryngeal muscles in respect to vowel devoicing in Japanese, *Phonetica*, **23**, 156-170.

Hirose, H. & Ushijima, T. (1978) Laryngeal control for voicing distinction in Japanese consonant production, *Phonetica*, **35**, 1-10.

Isshiki, N. & Ringel, R. (1964) Airflow during the production of selected consonants, *Journal of Speech and Hearing Research*, **7**, 233-244.

Jongman, A (1989) Duration of frication noise required for identification of English fricatives, *Journal of the Acoustical Society of America*, **85**, 1718-1725.

Keating, P. A. (1995) Effects of prosodic position on /t,d/ tongue/palate contact. In *Proceedings of the XIIIth international congress of phonetic sciences*, **3** (K. Elenius & P. Branderud, editors), pp. 432-435. Stockholm: Stockholm University.

Krakow, R. A. (1993) Nonsegmental influences on velum movement patterns: syllables, sentences, stress, and speaking rate. In *Nasals, nasalization, and the velum* (M. K. Huffman & R. A. Krakow, editors), pp. 87-116. San Diego: Academic Press.

Krakow, R. A., Bell-Berti, F. & Wang, Q. (1995) Supralaryngeal declination: evidence from the velum. In *Producing speech: contemporary issues: for Katherine Safford Harris* (F. Bell-Berti & L. Raphael, editors), pp. 333-353. Woodbury, NY: AIP Press.

Lass, R. (1984) *Phonology*. Cambridge: Cambridge University Press.

Laver, J. (1994) *Principles of Phonetics*. Cambridge: Cambridge University Press.

Lindblom, B. (1990) Explaining phonetic variation: a sketch of the H & H theory. In *Speech production and speech modelling* (W. J. Hardcastle & A. Marchal, editors), pp.403-439. Dordrecht: Kluwer.

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

Lisker, L. & Abramson, A. (1967) Some effects of context on voice onset time in English stops, *Language and Speech*, **10**, 1-28.

Lisker, L. Abramson, A., Cooper, F. & Schvey, M. (1969) Transillumination of the larynx in running speech, *Journal of the Acoustical Society of America*, **45**, 1544-1546.

Manuel, S. (1991) Some phonetic bases for the relative malleability of syllable-final versus syllable-initial consonants. In *Actes du XIIème congrès international des sciences phonétiques*, **5**, 118-121. Aix-en-Provence: Université de Provence.

Nespor, M. & Vogel, I. (1986) *Prosodic phonology*. Dordrecht: Foris.

Ohala, J. J. (1983) The origin of sound patterns in vocal tract constraints. In *The production of speech* (P. MacNeilage, editor), pp. 189-216. New York: Springer-Verlag.

Ohala, J. J. (1990) Respiratory activity in speech. In *Speech production and speech modelling* (W. J. Hardcastle & A. Marchal, editors), pp. 23-53. Dordrecht: Kluwer.

Pierrehumbert, J. & Talkin, D. (1992) Lenition of /h/ and glottal stop. In *Papers in laboratory phonology II: gesture, segment, prosody* (G. J. Docherty & D. R. Ladd, editors), pp. 90-117. Cambridge: Cambridge University Press.

Port, R. & Crawford, P. (1989) Incomplete neutralization and pragmatics in German, *Journal of Phonetics*, **17**, 257-282.

Port, R. & O'Dell, M. (1984) Neutralization of syllable-final voicing in German, *Journal of Phonetics*, **13**, 455-471.

Raphael, L. (1972) Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English, *Journal of the Acoustical Society of America*, **51**, 1296-1303.

Rothenberg, M. (1981) An interactive model for the voice source. In *Vocal fold physiology* (D. M. Bless & J. H. Abbs, editors), pp. 155-165. San Diego: College-Hill Press.

Scully, C. (1971) A comparison of /s/ and /z/ for an English speaker, *Language and Speech*, **14**, 187-200.

Soli, S. (1982) Structure and duration of vowels together specify fricative voicing, *Journal of the Acoustical Society of America*, **72**, 366-378.

Stevens, K., Blumstein, S., Glicksman, S., Burton, M. & Kurowski, K. (1992) Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters, *Journal of the Acoustical Society of America*, **91**, 2979-3000.

Vayra, M. & Fowler, C. (1992) Declination of supralaryngeal gestures in spoken Italian, *Phonetica*, **49**, 48-60.

Veatch, T. (1989) Word-final devoicing of fricatives in English. Paper presented at the Linguistic Society of America meeting, Washington, DC.

The Effect of Stress and Prosodic Phrasing on Duration, Acoustic Amplitude and Air Flow of Nasals in Estonian*

Matthew Gordon

1. Introduction

The acoustic and articulatory properties of segments may vary depending on prosodic position. In particular, there are two aspects of prosody which affect segments markedly: stress and prosodic phrasing. The effects of stress manifests themselves along many different acoustic and articulatory parameters: duration (e.g. Fry 1955), voice-onset-time (e.g. Cooper 1991), amplitude (e.g. Fry 1955), fundamental frequency (e.g. Fry 1958), tongue position (e.g. de Jong 1995) and jaw position (e.g. Beckman et al. 1992), velum height (e.g. Vaissière 1988), size of glottal opening (e.g. Cooper 1991), etc.

Less well-documented, but currently the subject of much research, is the effect of prosodic phrasing on segmental properties. By the term "prosodic phrasing" we mean the grouping of segments into progressively higher or larger prosodic constituents such as the syllable, the foot, the word, the phrase, the utterance, etc. (Selkirk 1984). Initial results of the research program examining the effects of prosodic phrasing on segments have yielded striking results, some of which are summarized below.

Probably the most familiar result concerns the lengthening effect seen in segments in the final position of prosodic domains (e.g. Oller 1973, Beckman et al. 1992, Wightman et al. 1992, etc.). Studies such as Wightman's indicate a general correlation between the level of the prosodic domain and the lengthening effect seen in final segments: the higher, or larger, the level, the greater the lengthening effect, in general. Lengthening is also observed in domain initial segments (Oller 1973). Another parameter along which segments vary as a function of prosodic phrasing is VOT. VOT has been shown to be longer for phrase initial stops than for phrase medial stops in English and Korean (Pierrehumbert and Talkin 1992 for English; Jun 1993 for Korean). In English, Cooper (1991) has shown that word initial voiceless stops have a longer VOT than word medial, but syllable initial, voiceless stops. He further demonstrates that the peak glottal opening for voiceless stops is greater for word initial than for word medial stops. Furthermore, there appears to be a cross-linguistic tendency for the amplitude of domain initial nasal consonants to be smaller the higher the domain (Fougeron 1996). Nasal flow experiments conducted by Fougeron and Keating (1996a) have also shown that prosodic phrasing affects the amount of nasal flow in French. The higher the prosodic domain, the less nasal flow there is in domain initial nasal consonants. The decrease in nasal flow, they argue, is largely the result of the velum being in higher position. The higher velum position results in a smaller velopharyngeal opening which allows for less air to escape through the nose. Thus, based on the results from the flow data, Fougeron and Keating infer that velum height is directly correlated with prosodic level: the higher the prosodic domain, the higher the velum position. Their results also show that the area of contact between the tongue and palate for the articulation of domain initial /t/ and /n/ in French (Fougeron and Keating 1996a) and /n/ in English (Fougeron and Keating 1996b) is larger the higher the prosodic domain.

The goal of the present study is to examine the acoustic and articulatory properties of syllable onset nasal stops in the initial position of various prosodic domains under differing stress conditions in Estonian. These four prosodic positions were considered under two stress conditions: the onset of stressed and unstressed syllables. The cross-classification of the four

---

* This paper is a written version of an ASA paper delivered in Indianapolis in May 1996 entitled "Nasal duration and amplitude as a function of stress and prosodic phrasing in Estonian." Thanks to participants in this experiment, and also to members of the UCLA Phonetics Laboratory and the ASA audience for suggestions and comments. Also, thanks to Cécile Fougeron, Chai-Shune Hsu and Ian Maddieson for proofreading this manuscript.

prosodic phrasing conditions by the two stress conditions yielded a total of eight different prosodic environments. The four prosodic domains examined were chosen, since they could easily be determined without a detailed study of Estonian prosodic phrasing. Utterance initial position corresponded to the commencement of speaking. Phrase-initial nasals followed a large intonation break after a vocative. Word initial nasals appeared at the beginning of orthographic (and also phonological) words, while syllable initial nasals were in syllable onset position in the middle of words.

2. Methods

Three experiments were performed, each designed to study a different acoustic or articulatory parameter: nasal flow, amplitude and duration. Four prosodic positions were examined: utterance initial position, phrase initial position, word initial position and syllable initial position. The speech of four native Estonian speakers (one woman and three men) was examined for the amplitude and duration experiments, three (one woman and two of three men examined in the amplitude and duration phases) for the nasal flow experiments involving prosodic phrasing, and two speakers (one woman and one man) for the nasal flow experiments involving stress.

The corpus for all three experiments consisted of a reiterant version of Estonian words. All of the targetted words were four syllables long, except for the one containing the syllable initial unstressed nasal which was three syllables long. In this case, the three syllable word was chosen, in order that the unstressed nasal would not be immediately preceded by a stressed vowel. The amplitude peak associated with primary stress in Estonian typically follows a stressed short vowel in an open syllable and falls on the onset of the following syllable. All syllables consisted of /n/ plus the vowel /a/. Subjects were primed using real words displaying similar stress patterns: e.g. [kátanaki] 'even as a crawfish net'; [alpákana] 'as an alpaca' and told to mimic the prosodic patterns of these words in their reiterant speech. The targetted words were placed in frame sentences consisting of real words. Between eight and ten repetitions of each sentence were recorded for all the experiments. The corpus appears in table 1; the underlined nasals were measured.

Table 1: Corpus

utterance initial, stressed:	Nánanana on sɤna.
utterance initial, unstressed:	Nanánana on sɤna.
	'N...... is a word.'
phrase initial, stressed:	Paul, nánanana on sɤna.
phrase initial, unstressed:	Paul, nanánana on sɤna.
	'Paul, n..... is a word.'
word initial, stressed:	See nánanana on sɤna.
word initial, unstressed:	See nanánana on sɤna.
	'This n..... is a word.'
syllable initial, stressed:	Nanánana on sɤna.
syllable initial, unstressed:	Nánana on sɤna.
	'N.... is a word.'

Primary stress in native Estonian words falls on the initial syllable. A large number of borrowed words, however, possess non-initial stress. For this reason, reiterant speech with non-initial stress were produced with relatively little difficulty for most speakers. Reiterant words with non-initial stress were not measured for speaker M2. As a result of this omission, only three of the eight conditions were examined for speaker M2: utterance initial stressed, phrase initial stressed, and word initial stressed. Syllable initial unstressed nasals were not examined, since they differed from the other examined nasals in both stress and phrasing, and hence would not have allowed for examinination of these factors independently of each other.

2.1 Nasal Flow

The first experiment examined peak nasal flow. Flow was recorded using a Rothenberg split mask. The raw flow data was converted to an amplitude display on Kay's Computerized Speech Lab from which measures of the peak amount of nasal flow during the nasals were made. Results of the effect of stress on nasal flow appear in figure 1 for the two speakers examined.

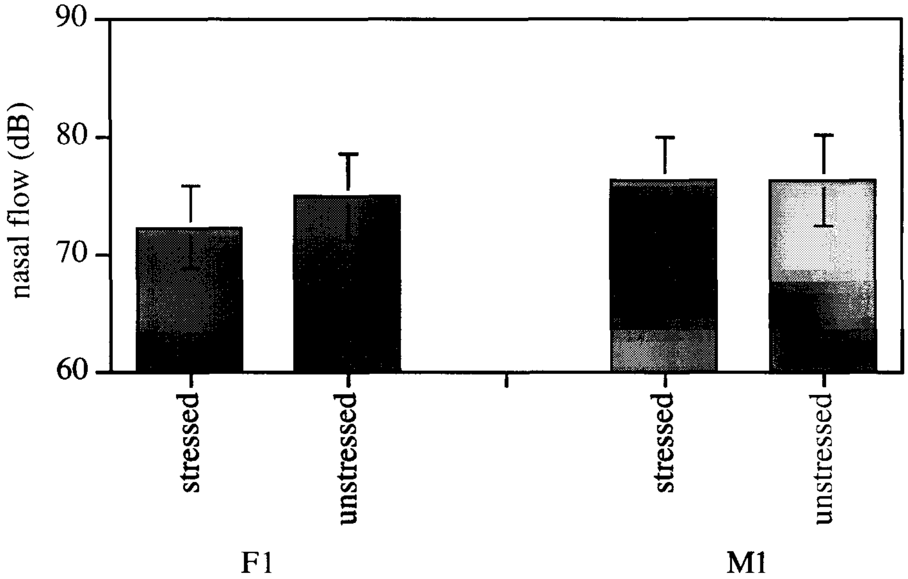


Figure 1: Peak nasal air flow as a function of stress

As is clear from figure 1, nasal flow only differs as a function of stress for speaker F1 but not for speaker M1. Speaker F1's stressed nasals are marked by less air flow than unstressed nasals. An ANOVA factorial revealed this difference to be highly significant [F (3, 69)=67.32, p<.0001]. In figure 2, the stress results are divided according to prosodic position.

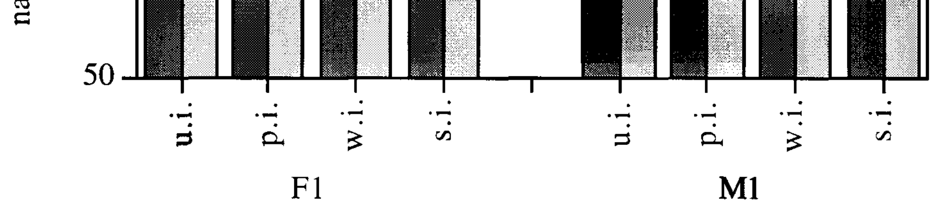


Figure 2: Peak nasal air flow as a function of stress by prosodic position

Nasal flow for speaker F1 differs significantly as a function of stress at all four prosodic levels. However, this difference goes in two different directions. At the highest three levels (utterance, phrase and word), stressed nasals have less air flow than unstressed nasals. This pattern, however, is reversed at the syllable level, where stressed nasals actually are characterized by significantly *greater* flow relative to their unstressed counterparts.

Speaker M1's nasals do not differ in a consistent direction according to prosodic phrasing. The only significant result (p=.0003) was for word-initial stressed nasals to be marked by less flow than word-initial unstressed nasals.

Next, we consider the effect of prosodic phrasing on nasal flow. Results from three speakers (F1, M1 and M2) are shown in figure 3, collapsing the two stress conditions. Syllable initial nasals were not examined for speaker M2 for reasons explained in section 2.

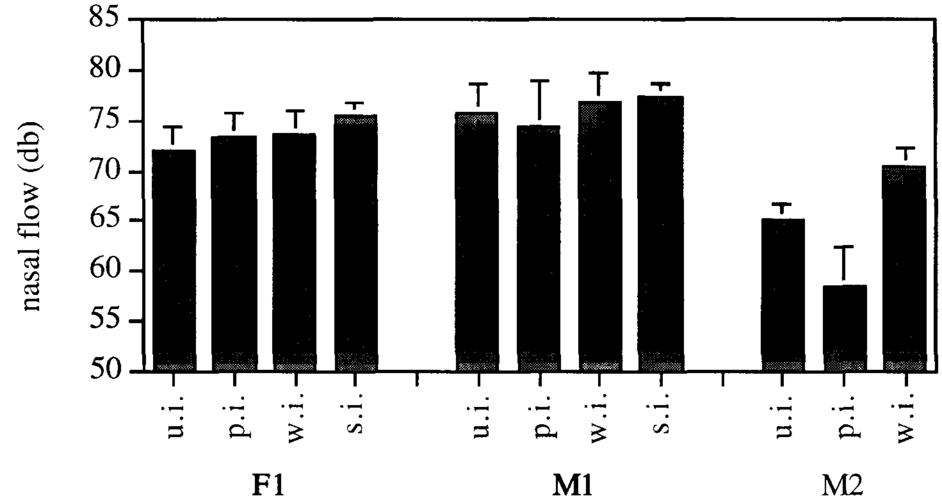


Figure 3: Peak nasal flow as a function of prosodic phrasing

If we first consider the three lowest prosodic groupings, the syllable, word and phrase a consistent trend is clear for both of the speakers for whom all levels were examined: the larger the prosodic constituent, the less the amount of air flow for domain-initial nasals. Speaker M2 also displays the same pattern for phrase and word initial nasals: phrase initial nasals have less flow than word initial nasals. All of differences involving the lower three levels of the hierarchy are statistically significant at the p<.05 level, except for phrase vs. word-initial for speaker F1 and word vs. syllable-initial for speaker M1.

Utterance-initial nasals, however, do not behave as might be expected given the position of the utterance at the top of the prosodic hierarchy. Only speaker F1's utterance initial nasals have the least amount of flow of the four phrasing conditions. Utterance-initial nasals for speakers M1 and M2 display greater flow than phrase-initial nasals. In the case of speaker M2, this difference is highly significant at the p<.01 level.

Looking back at figure 2, an interesting difference between speakers F1 and M1 is apparent. Nasal flow for speaker F1 hardly varies at all as a function of phrasing for unstressed nasals. All levels, however, are distinguished for stressed nasals. In contrast, in the case of speaker M1, greater differentiation is made in unstressed nasals. In fact, it is the unstressed nasals which are entirely responsible for the significant difference between phrase and word-initial nasals seen in figure 3.

2.2. Amplitude

Acoustic amplitude of nasals was measured in this experiment. The same corpus was recorded in a sound booth *after* the nasal flow data was recorded. Average RMS amplitude was calculated over the duration of the nasal as segmented from a waveform in conjunction with a spectrogram. In order to control for random fluctuations in speaking level, the average amplitude value for the nasal was subtracted from the average amplitude of the *first stressed* vowel in the examined word. The measured nasals and the first stressed vowels are shown in table 2. The relevant nasal is underlined; the relevant vowel is bold-faced.

Table 2: Corpus for amplitude experiment:
Amplitude of **vowel** minus amplitude of nasal:

utterance initial, stressed:	N**á**nanana on sɤna.
utterance initial, unstressed:	Nan**á**nana on sɤna.
phrase initial, stressed:	Paul, n**á**nanana on sɤna.
phrase initial, unstressed:	Paul, nan **á**nana on sɤna.
word initial, stressed:	See n**á**nanana on sɤna.
word initial, unstressed:	See nan **á**nana on sɤna.
syllable initial, stressed:	Nan**á**nana on sɤna.
syllable initial, unstressed:	N**á**nana on sɤna

The first stressed vowel and not the vowel immediately following the nasal was chosen because the vowel immediately following the targetted nasal was sometimes stressed and sometimes unstressed. Thus, any differences in relative nasal amplitude could be confounded by potential stress related differences in the amplitude of the reference vowel. Phrase-initial nasals were not examined for speaker M3. Syllable initial nasals could not be examined for speaker M2 for reasons elaborated earlier in section 2. The effect of prosodic phrasing on amplitude is shown in figure 4.

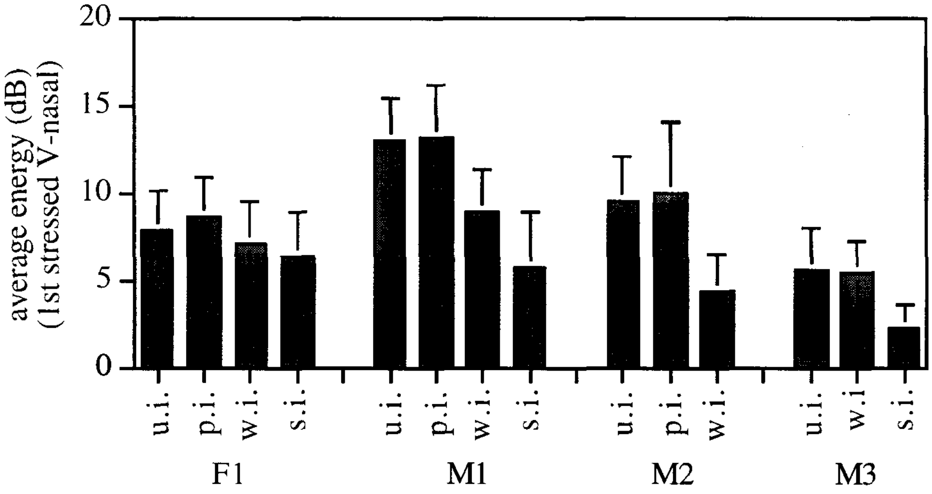


Figure 4: Amplitude difference between [n] and the 1st stressed vowel as a function of prosodic position

In figure 4, the higher the bar, the greater the difference in amplitude between the examined nasal and the first stressed vowel; i.e. the higher the bar, the less amplitude the nasal has relative to the vowel. All four speakers distinguish at least two levels in the prosodic hierarchy, and in the same direction: nasals in initial position of higher prosodic boundaries possess less amplitude relative to the first stressed vowel than nasals in initial position of lower prosodic boundaries. Phrase and word initial nasals were significantly different ($p<.05$) for all of the speakers for whom nasals in both of these positions were examined (speakers F1, M1 and M2). Phrase and syllable initial nasals were also differentiated ($p<.05$) for all speakers for whom the relevant nasals were examined (speakers F1 and M1). Speakers M1 and M3 also displayed a significant difference ($p<.05$) between word and syllable initial nasals. Speaker F1 showed a similar, but non-significant, trend for word initial nasals to have less amplitude than syllable initial nasals. It is interesting to note that none of the speakers distinguish utterance and phrase initial nasals in terms of amplitude.

The amplitude measures thus far are compatible with the results from the nasal flow experiment under the assumption that amplitude of nasal flow is correlated with the amount of air entering the nasal cavity. The greater the volume of air, the greater the amplitude of the nasal. Conversely, the less the volume of air, the smaller the amplitude of the nasal. The predicted trend for amplitude and nasal flow to be correlated appears to be corroborated for the most part by measurements of nasal flow and amplitude for individual speakers (cf. figures 3 and 4): where the difference in amplitude between the nasal and the vowel is largest, flow is greatest, and vice versa. For example, nasal flow and amplitude are smallest for speaker M1's phrase-initial nasals, while nasal flow and amplitude are greatest for the same speaker's syllable initial nasals. The correlation also holds for speaker M2 as well, though the correlation is not as clear for speaker F1. Although her phrase initial nasals have the least amplitude, they do not possess the least flow. In general, however, figures 3 and 4 suggest that flow and amplitude are at least partially predictable from each other.

The link between flow and amplitude is not as convincing, however, when one compares figure 1 with results for the phase of the experiment examining the effect of stress on amplitude, depicted in figure 5. (Recall that speaker M2 was omitted from the stress comparison.)

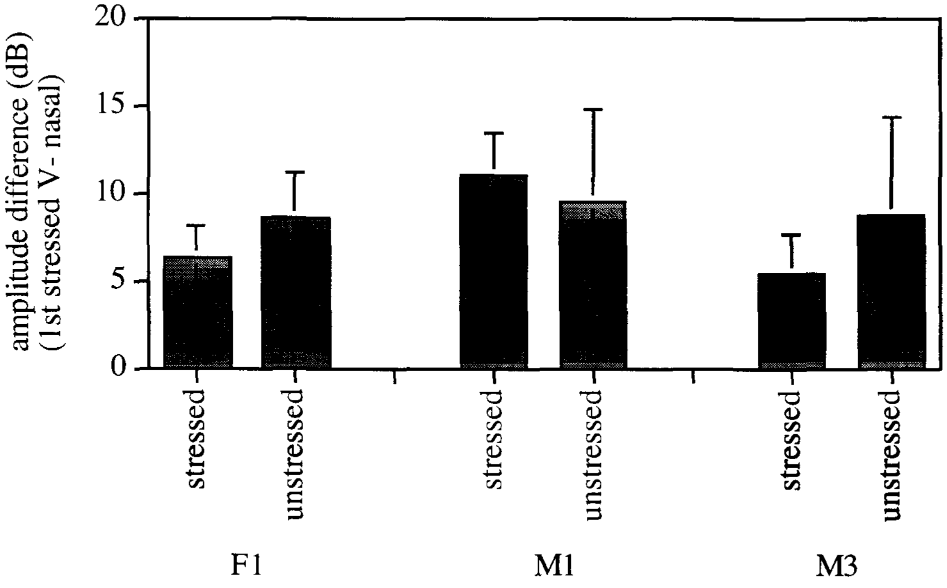


Figure 5: Amplitude difference between [n] and the 1st stressed vowel as a function of prosodic position

Stress was shown by a two-factor ANOVA (stress and prosodic phrase position) to have a significant effect on nasal amplitude for two of the three speakers. For speakers F1 and M3, nasals in the onset of stressed syllables have significantly ($p<.01$) greater amplitude relative to the first stressed vowel than nasals in the onset of unstressed syllables. Speaker M1 does not differentiate stressed and unstressed nasals in terms of amplitude.

The results in figure 5 do not follow the predicted pattern given the results in figure 1, which showed that stress was associated with decreased nasal air flow for speaker F1. One might have expected that the decrease in air flow would trigger a decrease in amplitude. This apparent contradiction is presumably due to a difference in air flow through the glottis between nasals in stressed and unstressed positions. An increase in air flow associated with stress would presumably raise the overall amplitude of the nasal. Because the amplitude measure encompasses acoustic energy of the nasal across all frequencies, the energy contributed by the passage of air through the glottis is not differentiated from the energy contributed by the nasal cavity. Thus, it seems plausible that the discrepancy between amplitude in stressed and unstressed nasals is due to a difference in the amount of transglottal air flow rather than a difference in the amount of air flow through the nasal cavity. A measure of energy differentiated by frequency would perhaps shed light on the matter.

2.3. Duration

The next experiment examined the effect of stress and phrasal position on nasal duration. Duration was measured from a waveform with reference to a spectrogram. First, the effect of stress on duration is considered in figure 6.

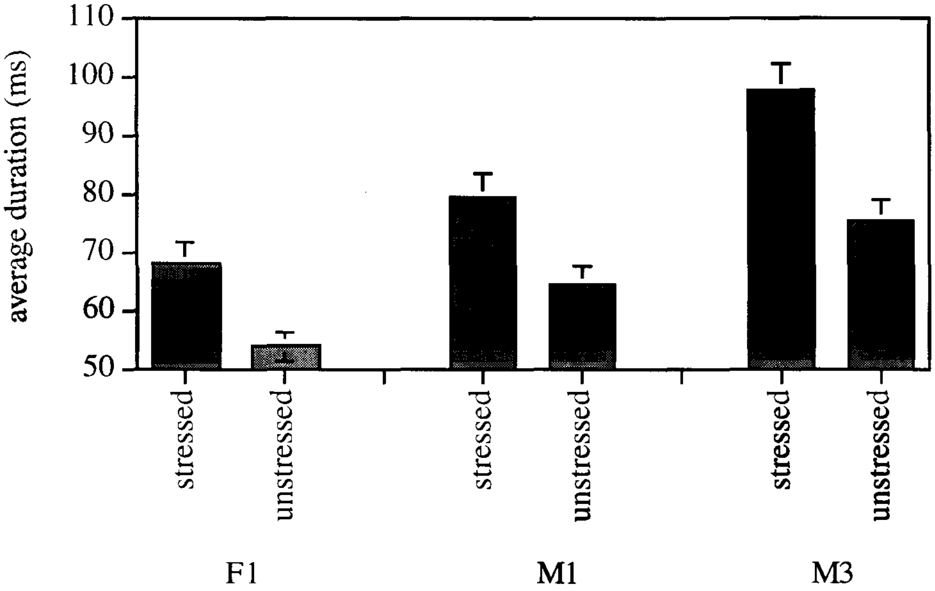


Figure 6: Nasal duration as a function of stress

The result is clear for all three speakers examined. Nasals in the onset of stressed syllables are significantly longer than nasals in the onset of unstressed syllables, duplicating earlier results (Lehiste 1966; Gordon 1995). The effect of phrase level on duration is far less clear, as shown in figure 7.

For two speakers (M2 and M3), there is a consistent pattern of decreased duration, the smaller the prosodic domain. Note, however, given the small number of tokens, none of these differences fall under the $p<.05$ level of significant, although the difference between utterance initial and word-initial nasals for speaker M2 and the difference between utterance initial and syllable initial nasals for speaker M3 approach significance: $p<.0757$ and $p<.0715$, respectively. None of the differences for speaker F1 are significant. The strongest results come from speaker M1, for whom the size of the prosodic domain is *inversely* correlated with nasal duration. The difference between utterance initial and syllable initial and between phrase initial and syllable initial nasals are highly significant at the $p<.01$ level for speaker M1, while all other differences are nearly significant ($p<.10$), except for the difference between utterance and phrase initial nasals.

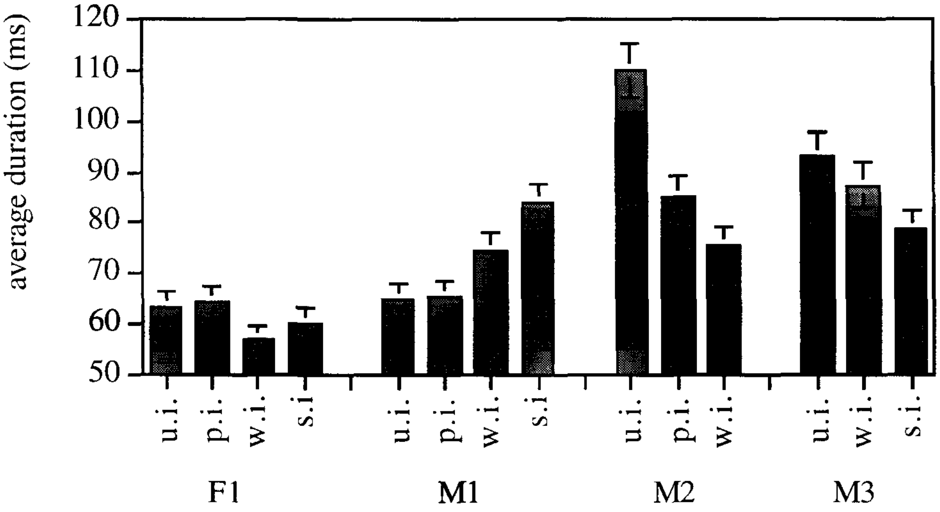


Figure 7: Nasal duration as a function of prosodic level

3. Summary and Discussion

Measurements of duration, flow and amplitude of nasals in Estonian suggest a general trend for nasals to be "strengthened" both under stress and following prosodic junctures, where the degree of strengthening is correlated with the size of the juncture. This strengthening is manifested either as lengthening of the nasal, or more extreme raising of the velum.

First, we summarize the effect of stress on nasals. The acoustic parameter on which stress seems to exert the most marked and consistent influence across speakers is duration. Nasals in the onset of stressed syllables were significantly longer for all speakers examined. Stress also influenced nasal flow and amplitude, though the results were not as consistent as the duration result. One speaker showed decreased nasal flow in nasals in the onset of stressed syllables relative to nasals in the onset of unstressed syllables, suggesting a heightened velum position in stressed syllables, following Fougeron and Keating's reasoning (1996a). The other speaker's stressed nasals, however, were not marked by decreased air flow relative to his unstressed nasals.

The effect of stress on nasal amplitude is speaker dependent. Stressed nasals have greater amplitude than unstressed nasals for two of the three speakers examined. The other speaker shows no significant difference in amplitude based on stress.

Those parameters which were least reliably correlated with stress, nasal flow and acoustic amplitude, turned out to be precisely the properties most consistently associated with differences in prosodic level. With the exception of utterance initial nasals, domain initial nasals had lesser flow, the higher the domain. While nasal air flow for all speakers did not distinguish every prosodic level on a consistent basis, all speakers did distinguish at least two levels, and always in the predicted direction. Acoustic amplitude of domain initial nasals for all speakers is also correlated with the size of the domain: the larger the domain the greater the difference in amplitude between the nasal and the first stressed vowel of the examined word. In other words, the larger the domain, the less the amplitude of the nasal.

Interestingly, the property most closely linked to stress, duration, was the property which least reliably differentiated prosodic levels. Two speakers (M2 and M3) tended to lengthen nasals at the boundary of higher domains, while one speaker (M1) did just the opposite and shortened nasals at the boundary of higher domains. Yet another speaker did not display a consistent correlation between prosodic level and duration.

The tendency for a phonetic property to be consistently correlated with either stress or phrasing level but not both suggests that functional load plays a role in the exploitation of various phonetic parameters to signal stress or phrasing level. Where a property is closely involved with the demarcation of either stress or phrasing level, it will be less available for serving other functions. Thus, because onset duration carries such an important load in the marking of stress in Estonian, it is less available for signaling phrase boundaries. Conversely, velum height and amplitude is used to distinguish prosodic levels, and is thus not as readily available for manipulation in cueing stress. Assuming that avoidance of functional overload is at work in Estonian, as well as in other languages, then we would expect that a language with different properties associated with stress might also display differences in the properties associated with phrase level. For example, in a language which does not signal stress by increasing the duration of the onset might be expected to vary duration more as a function of phrasing level. However, while the functional load explanation may offer some answers to why certain properties but not others are manipulated as a function of phrasing or stress, such an explanation does not account for all the data across languages. For example, VOT is correlated with stress in English (Cooper 1991), yet it also varies as a function of the size of the prosodic domain in which it occurs in initial position. More data from other languages is necessary to examine the extent to which languages try to avoid using a single parameter to signal both stress and prosodic level.

References

Beckman, M., Edwards, J., and Fletcher J. 1992. "Prosodic structure and tempo in a sonority model of articulatory dynamics," in *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, edited by G. Docherty and D. R. Ladd. Cambridge U. Press, Cambridge.

Cooper, A. 1991. "Laryngeal and oral gestures in English /p, t, k/", *Proceedings of the XIIth International Congress of Phonetic Sciences*, 1991, Aix-en-Provence, France, vol.2, 50-53.

de Jong, K. 1995. "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," *JASA* 97, 491-504.

Fougeron C. 1996 "Variation de débit nasal en fonction de la position prosodique de [n] et [ã] en français", in *Actes des XXIè J.E.P.*, Avignon, France.

Fougeron, C. and Keating, P. 1996a "The influence of prosodic position on velic and lingual articulations in French: Evidence from EPG and airflow data," this volume.

Fougeron, C. and Keating, P. 1996b. "Articulatory strengthening in prosodic domain-initial position," this volume.

Fry, D.B. 1955 "Duration and intensity as physical correlates of linguistic stress," *JASA* 27, 1765-8.

Fry, D.B. 1958. "Experiments in the perception of stress," *Language and Speech* 1, 126-152.

Gordon, M. 1995. "Acoustic properties of primary and secondary word-level stress in Estonian," poster presented at 130th meeting of the Acoustical Society of America, St. Louis, MO, November 1995.

Jun, S. 1993. *The phonetics and phonology of Korean prosody*. PhD dissertation, Ohio State U.

Lehiste, I. 1966. *Consonant quantity and phonological units in Estonian*. Indiana University Press: Bloomington.

Oller, D. K. 1973. "The effect of position in utterance on speech segment duration in English," *JASA* 54, 1235-47.

Pierrehumbert, J. and Talkin, D. 1992. "Lenition of /h/ and glottal stop," in *Papers in Laboratory Phonology II*, edited by G. Docherty and D.R. Ladd. Cambridge U. Press, Cambridge.

Selkirk, E. 1984. *Phonology and syntax: the relation between sound and structure*. MIT Press, Cambridge, MA.

Vaissière, J. 1988. "Prediction of velum movement from phonological specifications", *Phonetica* 45, 122-139.

Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P.J. 1992. "Segmental durations in the vicinity of prosodic phrase boundaries," *JASA* 92, 1707-17.

Phonetic Universals*

Ian Maddieson

1. Introduction

There are two general approaches to the question of universals in phonetics. One approach is based on drawing a distinction between the learned and the automatic or 'mechanical' aspects of speech behaviour. The second on considering the way that the phonetic aspects of language must be structured in order to satisfy the functional demands of communication. For convenience, these two approaches may be distinguished as the 'mechanistic' and the 'ecological', but they are not as disjunct as the labels attached to them may suggest. Nonetheless, because research efforts tend to be directed more toward one or the other there are two rather separate literatures on them. This chapter will be structured as a discussion of these two approaches in turn.

2. Learned versus inherent patterns

From the mechanistic viewpoint the particular way a given human language sounds is due to the interaction of two classes of processes. One class of processes is responsible for language-specific details, according to phonetic regularities governing the relationship between the phonological structure of the language and linguistically-relevant physical phonetic parameters. The other class shapes the sound patterns in ways that are necessarily universal, that is, they are the result of inherent properties of the mechanisms by which speech is produced and processed, and of the general physical laws under which the universe operates. The phonological structures and the language-specific rules must be learned as part of the process of acquiring a command of the individual language concerned. By contrast, it would be meaningless to talk of a speaker learning the processes that shape universals since these are simply part of the context within which speech exists.

As an example of the kind of distinction being drawn, consider a pair of English words such as 'pill' and 'kill'. The consonants at the beginnings of these words are usually pronounced with a substantial amount of aspiration, slightly longer with /k/ than with /p/. The fact that aspiration occurs here is a language-specific property of English; this can be expressed by saying that English speakers must learn a rule specifying that voiceless plosives are aspirated when they are initial in a stressed syllable. Other languages have a different distribution of aspiration, or lack it altogether. In contrast to this language-specific pattern, the fact that the duration of the aspiration is on average longer with the velar plosive than the bilabial is usually attributed to factors that are inherent in the use of this place contrast. This is because parallel variations in the duration of the voice onset interval can be found in languages that have no aspiration, such as Spanish and Hungarian, or have contrasting aspirated and unaspirated voiceless plosives, such as Cantonese and Hindi (Lisker and Abramson 1964).

As in this example, much of the work on 'mechanistic' universals has been based on the simple observation of patterns of repeated similarity across languages. When a pattern is seen to be widely repeated, the assumption is often made that it must in some way be a consequence of the fact that human beings all make use of the same basic equipment for speaking and listening

---

* To appear in the *Handbook of Phonetic Sciences* edited by John Laver and William J. Hardcastle, Blackwells Publishers, Oxford (1996).

and are subject to the same physical laws. Subsequently, a search may be conducted for an explanation of the pattern in these terms. One result of this mode of working, as well as the lack of precision of many of the existing models of speech processes, is that there are a good many putative universals for which no accepted explanation is available. As our understanding of speech production and perception mechanisms grows, it is increasingly likely that genuinely universal patterns can be predicted from more precise models of the processes involved. However, at the present time competing explanations have been proposed for many putative universals. For as long as this remains true, some uncertainty must be associated with the claim that a given pattern is in fact appropriately considered universal.

We can illustrate this point with the commonly observed place-related differences in voice onset time mentioned above in connection with English 'kill' and 'pill'. One type of explanation focusses on the velocity of movement of the articulators concerned. The releasing movement of a velar closure is slower than that for a bilabial closure, in part because the rotational movement of the jaw accelerates the movement of the lower lip in comparison with the back of the tongue (Kuehn and Moll 1976). The distance from the pivot point of the rotation is greater for the lip than for the tongue back. Hence for a given amount of angular motion of the jaw the lower lip will move further in the same time than a point on the tongue does, as shown schematically in figure 1. (This is probably not the only factor involved since differences between the two articulators in their mass and compliance may also contribute to a faster labial than velar release, as will be discussed in later section.) A different articulator speed affects voice onset time in the following way: Because the oral aperture through which air is escaping increases at a slower rate after a velar stop release, a longer time elapses before the crucial transglottal pressure difference required to initiate voicing is attained. This proposed explanation appeals to the mechanical constraints on the jaw, general laws of fluid dynamics, and the physiology of the vocal folds — clearly universal properties.

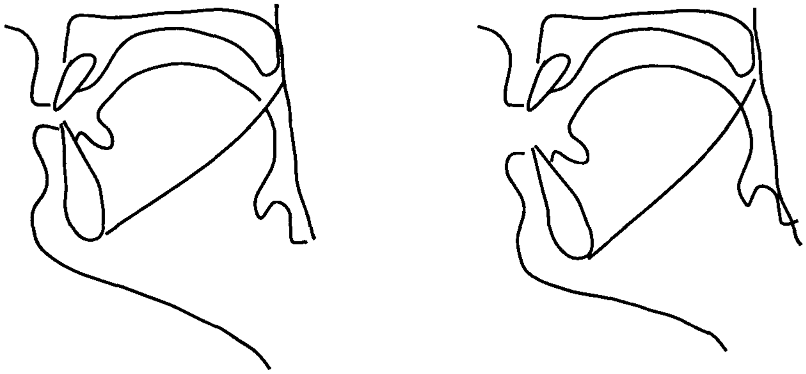


Figure 1. Schematic representation of the effect of jaw rotation. A 10° shift in jaw angle moves the lips apart more than the tongue back and velum.

An alternative account would explain the voice onset difference between 'pill' and 'kill' as resulting from different closure durations for stops at different places of articulation (itself a putative universal, which will be discussed further below), coupled with a stable duration for the vocal fold abduction gesture. This stable duration would be a learned feature of English voiceless stops in this particular position. Weismer (1980) measured the mean duration of the entire voiceless interval (from offset of voicing at the formation of the stop closure to onset of voicing after the release and the aspiration interval) for English words with initial /p/ and /k/, spoken in a carrier phrase, and found it to be the same in these two cases. This, as well as other evidence cited by Weismer, suggests that there is an abduction-adduction cycle of the vocal cords for voiceless stops which is longer in duration than the closure and has a constant time course, anchored to the onset of closure. If this is correct, then rate of articulator movement may have nothing to do with the different aspiration durations. A shorter hold of the closure for a velar stop than for a bilabial stop will automatically increase the time interval of the following aspiration by the amount of the difference, as is schematically shown in Figure 2. Such a precise trade-off between component durations is suggested by data on English from two classic studies. Umeda (1977) measured the mean closure duration of /k/ as 20 ms shorter than the closure of /p/ (69 vs 89 ms); Lisker and Abramson (1964) measured the aspiration duration for /k/ as 22 ms longer than that for /p/ (80 vs 58 ms).

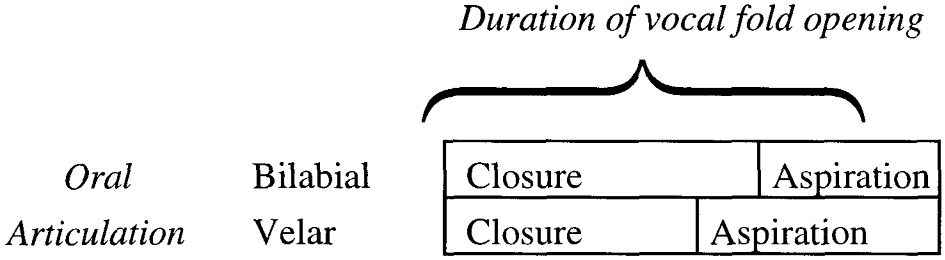


Figure 2. Schematic representation of place differences in aspirated stops from constant vocal cord abduction plus different closure duration.

When these competing accounts are considered together, it appears that there may not be a single universal accounting for a relationship between place of articulation and onset of voicing. The first explanation predicts a difference between bilabials and velars only if the vocal folds are together at, or very soon after, the release of the oral occlusion, and not if the vocal folds are still in an open position at the release. Note that if the vocal folds are not brought together until some 50 ms or more after the velar closure for /k/ has been released, it is unlikely that supraglottal pressure is high enough to delay the initiation of voicing. Intraoral pressure drops very rapidly once the closure is broken, and the tongue dorsum can be expected to have lowered by some 4 to 5 mm by 50 ms after release. Weismer's proposal for fixed vocal fold opening duration fits the situation better. The case is different for /b/ and /g/ because the vocal folds are adducted in these consonants; the explanation offered by Weismer does not apply. The fact that Lisker and Abramson (1964) found a similar difference of about 20 ms in the voice onset time between both the English consonant pairs /b/ ~ /g/ and /p/ ~ /k/ may therefore be coincidental. These two differences may be the result of different processes which in English operate in disjunct environments, but produce superficially similar outcomes.

We therefore stress that observing a pattern of cross-language similarities is only a very preliminary step. Understanding the basis on which the pattern is founded is the primary goal. What seems to be a unified pattern may have several causes, and conversely a single effect may have different outcomes depending on variations in other factors. Finding an apparent counterexample to a proposed universal may not invalidate its status as a universal unless all relevant factors are constant. And a pattern that is 'mechanical' in one context may be generalized to others and thus become part of what a speaker must learn about a language.

In the next section, a brief catalogue of some of the most widely suggested 'mechanistic' universals will be provided. In each case a very short discussion of proposed factors on which the universal pattern might depend will be provided together with references to some of the most important work on the topic.

3. A brief catalogue of proposed mechanistic phonetic universals

3.1 Universals relating to vowel height.

Several universals relating to vowel height have been proposed. Other things being equal, higher vowels have higher F0 than lower vowels. This correlation, sometimes referred to as the intrinsic pitch of vowels, has been documented in a large number of languages (Whalen and Levitt 1995). We add here data from Iaai, using an equal number of vowels in the high, mid and low classes. F0 was measured at the mid-point of vowels in words spoken in isolation. These results are given in Table 1.

Table 1. Vowel Height and F0 in Iaai (values rounded to nearest integer).

	Women (3 speakers)			**Men (2 speakers)**		
	n	***F0***	***s.d.***	***n***	***F0***	***s.d.***
High **[i, y, u]**	43	240	22	37	152	15
Mid **[e, ɤ, o]**	41	233	15	40	144	12
Low **[æ, a, ɔ]**	37	224	19	37	136	13

One class of explanations suggests that the intrinsic pitch effect arises because of a mechanical linkage between the supralaryngeal and laryngeal tissues, resulting in increased tension on the vocal folds as the tongue is raised (Ohala 1978). However, there are puzzling discrepancies between results obtained when different regions of the vowel space are examined, and some evidence that the level of activity of the cricothyroid muscle — the muscle most involved in the active control of F0 differences — is itself positively correlated with vowel height (Honda and Fujimura 1991). This suggests that the intrinsic pitch effect may involve some active targeting of different F0 values for different vowels. One possibility is that in a high vowel the (low) F1 is close enough to the F0 value to produce a perceptual integration of the two, usually believed to occur when two spectral components are closer together than 3.5 Bark (see xxxxxx in this volume). Since high vowels appear higher in pitch, speakers might then mimic this effect in their production. Another suggestion is that the intrinsic pitch effect is a consequence of the perceptual reliance on a speaker's characteristic F0 as a calibration tool for interpreting the speaker-dependent variation in F1 and other formants (Hoemeke and Diehl 1994). On this view, if a high vowel is produced with a higher F0, it triggers an interpretation that the vowel is higher than would otherwise be the case, reinforcing the percept of its height. (The calibration works as follows: A high mean F0 indicates that a speaker has a small vocal tract and all formants will be on the high side of average; they must therefore be equated with lower values from larger speakers. The imputed lower F1 correlates with a perceptually higher vowel.)

Other things being equal, higher vowels are shorter than lower vowels. This observation, like most others concerning segment durations, is traditionally based on measures of the intervals between acoustic landmarks. The effect, sometimes known as intrinsic vowel duration, is usually explained as due to the fact that a greater distance must be moved by the tongue and jaw to go from a consonantal constriction to a lower vowel and back again than is required for a higher vowel (Catford 1977). There are few studies that address this point, although Lindblom (1967) showed that short and long Swedish /ɑ/ vowels take longer than the /i/ vowels of the corresponding length category and that differences in the magnitude of mandibular movement are implicated in this difference. However, the magnitude of the vowel durational difference was actually less than might have been expected from the differences in movement. In his experiment, the vowels were surrounded by bilabial consonants, and some reorganization of the coordination between lip and jaw movements seems to be involved in limiting the durational differences.

Other things being equal, higher vowels also have a greater tendency to devoicing than lower vowels. Jaeger (1978) catalogued a number of cases from language descriptions where either only high vowels were devoiced in a particular environment, or higher vowels showed a greater tendency to occur voiceless than lower ones. This pattern is attributable to aerodynamic conditions: since the higher the vowel is the more constricted the oral passage is, a higher vowel creates greater impedance to oral airflow and therefore tends to raise supraglottal pressure. Thus over a certain range of laryngeal settings the pressure differential across the glottis in a high vowel will fall below the critical level required to initiate or sustain voicing, whereas a lower vowel would permit the voicing to occur.

3.2 Universals related to consonant voicing.

A number of other proposed universals relate to the voicing state of consonants. Most often these are discussed in connection with obstruents, as these are the consonants that most often provide an opportunity to contrast voiced and voiceless cognates. Other things being equal, the vowel before a voiced consonant is longer than that before its voiceless counterpart. For example, in English, the vowel in 'bid' is longer than that in 'bit'. It is often remarked that the difference is greater in English than in other languages, but some difference is reported in a wide variety of languages, and not just in closed syllables (as the Italian data below illustrates). The general effect is often explained as a consequence of the need to make a more forceful gesture of closure for a voiceless consonant, since the unimpeded flow of air through an open glottis provides greater resistance to the formation of an oral seal (Chen 1970). The resulting more forceful gesture results in a faster movement of the articulators and a more rapid achievement of the consonantal target position. There are several problems with this explanation. First of all, studies of articulatory movements show that the *onset* of the closing movement occurs earlier when the consonant is voiceless, so that more than a faster rate of closure is involved. We will return to this point later. Secondly, sonorant consonants in the coda are also affected; thus in a pair of words such as 'build' and 'built' both components of the [ɪl] sequence differ in length, and this effect extends to nasals in pairs such as 'send' and 'sent'. Raphael et al (1975) found that the length difference in the nasals was actually twice as great as that in the vowels in this pair. The articulatory closure is made as the nasal is initiated, and since the nasal itself is voiced in both environments there is unlikely to be any great difference in the resistance to closure between these two cases.

Other phoneticians have suggested that this type of durational difference is due to a perceptual mechanism. One proposal is that the continuity of voicing from a vowel into a voiced obstruent might make it harder to determine the end of the vowel, creating an "auditory illusion" of greater length. Consequently, speakers internalize a longer vowel target in such words (Javkin 1976). Another interpretation is that the vowel length difference is due to exploitation of a perceptual contrast effect to signal the voicing distinction. This idea relates the vowel length difference to another observational universal, namely that the acoustic duration of voiced consonants, especially obstruents, is shorter than that of corresponding voiceless ones. A longer preceding vowel makes the duration of a (short) voiced consonant appear even shorter in much the same way as, in the visual domain, changing the brightness of a background colour can change the perception of the brightness of an item in the foregound (Kluender, Diehl and Wright 1988; but see Fowler 1992). In this way the salience of the voicing distinction might be enhanced.

It is possible that the difference in consonant durations may account for the vowel duration pattern without appealing to enhancement of a perceptual contrast. Suppose that in a CVCV structure a normal production strategy is to control the duration of some longer unit, for example by fixing the phasing between the acoustic onsets of the vowels. With this requirement, a longer intervening consonant must "borrow" duration from the preceding vowel. An explanation of this general kind seems to account quite well for the shorter vowels that occur before geminate consonants in Italian in comparison with their singleton counterparts (Smith 1995). An unpublished study by Dunn (1993) suggests that taking the consonant duration as primary also provides a good account of the differences attributable to voicing in Italian. Some of her results comparing intervocalic /p/ and /b/ production and the preceding vowel in the nonsense words 'tapa' and 'taba' are shown in Table 2. Dunn used an LED tracking system to follow the lip movements in these sequences and also made the conventional measurements of acoustic durations. Reading down the rows of Table 2, we see that mean acoustic vowel duration (from the burst for /t/ to the labial closure) is on the order of 15 ms shorter before /p/ than before /b/. This matches very closely with the difference in the latency of the onset of the articulatory movement of the lips toward closure, i.e. the time between the /t/ burst and the beginning of the labial closing gesture. In other words, the lips start to close about 15 ms sooner to form a /p/. Once closed, they stay together longer. In fact, the /p/ - /b/ closure duration difference is longer than the difference in the preceding vowel. Although the closing velocity is indeed faster in forming a /p/, the duration of the closing movement is the same for /p/ and /b/, suggesting that the movement rate is not affecting the preceding segment durations, but is rather part of a strategy to maintain the longer closure for the voiceless consonant by making a more forceful closure. The peak articulatory position is held longer for /p/ than for /b/, especially by speaker B. Finally, the interval from the articulatory onset of the first /a/ vowel to the onset of the second is shown. This is longer with /p/ than with /b/, suggesting that /p/ also "borrows" some duration from the second vowel, particularly again for speaker B. Adding together the earlier onset of the closing movement for an intervocalic /p/ and the later onset of the movement toward the second vowel after /p/ we obtain a close match with the added acoustic duration of /p/ compared with /b/.

Table 2. Acoustic and articulatory measures of intervocalic /p/, /b/ for two Italian speakers (after Dunn 1993). All values are means of 20 repetitions.

	Speaker A			**Speaker B**		
	tapa	**tab**a	p/b diff	**tap**a	**tab**a	p/b diff
Acoustic vowel duration (ms)	157	173	- 16	191	208	- 17
Latency of movement onset for consonant (ms)	100	115	- 15	131	145	- 14
Acoustic consonant duration (ms)	99	76	+ 23	118	86	+ 32
Peak velocity of consonant closure (mm/sec)	275	210	+ 65	335	305	+ 30
Duration of onset movement (ms)	67	65	+ 2	78	79	- 1
Held duration of articulatory peak (ms)	60	53	+ 7	73	51	+ 22
V-onset to V-onset interval (ms)	256	249	+ 7	309	294	+ 15

It is thus possible to view the voicing-dependent differences in vowel duration as due to the interaction of larger-scale timing principles which function to create similar durations for similarly-structured words and a voicing-dependent effect on the inherent duration of the closure of stops. Timing principles of this type, though not universal, are common. The burden of explanation is therefore shifted to the consonant duration difference; a reasonable proposal is that this has an aerodynamic basis. Vocal fold vibration cannot be sustained when there is high resistance to airflow above the glottis, as in the case of a stop. A variety of manoevres to expand the size of the oral cavity, such as lowering the larynx, can reduce the supraglottal resistance. Similarly, a shortened duration of closure and a less extensive contact area both help to increase the likelihood that vocal fold vibration will continue through the closure until the moment of release (The shortened duration just limits the length of time that voicing needs to be sustained, whereas a less extensive contact leaves more surface area to absorb rising intraoral pressure). These patterns are not mechanical *consequences* of producing voicing, but are good ways of using vocal tract mechanics given a goal of realizing voicing in a stop consonant.

Another familiar observational universal is the tendency for the fundamental frequency (F0) to be higher after a voiceless consonant than after a voiced one (other factors affecting F0 being equal). This observation has been much discussed, and like other universals, interpreted in various ways. For many, it is viewed solely as the consequence of "the lowering effect of voiced consonants" (Mohr 1971). But there are probably (at least) two separate effects, one a lowering process and another which is responsible for raising pitch after voiceless consonants. The lowering is sometimes argued to be due not to the fact of voicing alone but to a process that lowers the F0 when obstruency and voicing are combined (Hombert 1978). For example, as noted above, keeping the vocal folds vibrating when there is a closure in the oral cavity is problematical. It can be assisted by lowering the larynx, and this lowering might incidentally mimic some aspect of the active pitch control regimen. After all, there is known to be an association between the height of the larynx, usually measured by the height of the thyroid notch, and active control of pitch (Ewan and Krones 1974, Riordan 1980). Ohala (1980) reasoned that if this was the explanation, then no lowering should be found following ordinary voiced nasal consonants as these allow a relatively high volume of air to continue to flow through the glottis, since the nasal passage is open. As predicted, she found that in the Hindi speaker she studied voiced stops did have a lowered F0, producing a rising F0 onset to the following vowel, but after nasals the F0 trajectory was flat. However, Maddieson (1984b) found that in Burmese, voiced nasals and laterals had a rising onset like that of voiced stops. Bickley and Stevens (1986, 1991) showed that vocal fold vibration drops in frequency, presumably due to aerodynamic conditions, when the oral escape channel is constricted but not closed by an experimenter-controlled manipulation. If the escape is sufficiently narrowed in the nasals and laterals of certain languages this may account for lowering sometimes being observed with consonants of these types.

The link between voicing and lowered F0 has received much more attention than its counterpart - the association between voicelessness and raised F0. In fact, in most studies which have measured both, the amount that F0 is raised after a voiceless obstruent is considerably greater than the amount of lowering seen after a voiced obstruent (comparing values at the vowel onset to a target F0 later in the following vowel). Data from a male speaker of Thai, based on Gandour and Maddieson (1976), are shown in Figure 3. As this is a tonal language, with a high functional load for tones, there is good reason for taking the target pitch on the following vowel to be relatively controlled. Yet mean onset of F0 after /p/ is about 7 Hz higher than after /b/, and the decline is sharper after /p/ than is the rise after /b/. Informal suggestions have been made that raised F0 occurs because of high transglottal air flow when a stop or other obstruent constriction is released while the vocal folds still remain apart. Since an increased flow — such as that resulting from an unexpected push on the chest while speaking — does increase F0 (Ladefoged

1967, Sonesson 1982), the raised F0 may be the result of the beginning of the vowel being produced while the vocal folds are not yet fully adducted. A slightly wider aperture between the folds would allow greater flow but the folds would still be close enough to vibrate. In this case it might be anticipated that voiceless aspirated stops would show greater F0 elevation than voiceless unspirated ones. As the Thai data in Figure 3 show, this is far from a universal finding, so other factors must also be involved.

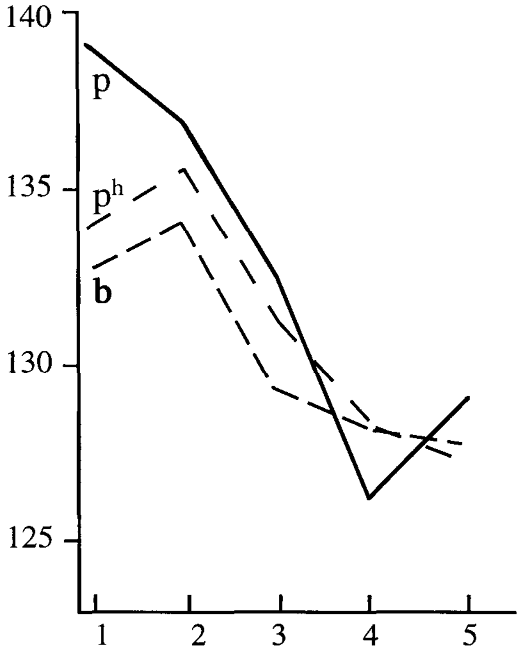


Figure 3. Mean F0 contours in Hz after consonants differing in voicing, averaged across the five tonal categories of Thai (measurements taken at 5 equally spaced intervals through the duration of the vowel).

An alternative theory is based on the observation that the cricothyroid muscles can be actively involved in the production of voiceless consonants. Contraction of these muscles increases the distance between the attachments of the vocal folds, thereby stretching the folds longitudinally and increasing the frequency of their vibration when other conditions are appropriate for voicing to occur (Sonesson 1982). Cricothyroid activity is thus the principal way of increasing F0 for intonational peaks, high tones and other upward pitch excursions. In a study of three subjects (two speakers of American English and one of Dutch), Löfqvist, Baer, McGarr and Seider Story (1989) showed that the level of electrical activity in the cricothyroid muscles is also higher during voiceless stops, affricates and fricatives than during their voiced counterparts. Moreover, since the peak of this activity occurs relatively early in the acoustic duration of the voiceless consonants, Löfqvist et al. argue that it is directly associated with their voicelessness, and that the tensing of the vocal folds assists in inhibiting vocal fold vibrations. The raising of F0 is incidental; because the relaxation time of these muscles is relatively long the folds are still somewhat tensed at the vowel onset following voiceless consonants. A persuasive detail is that the speaker and consonant class with the smallest cricothyroid activity difference (the affricates of one of the English speakers) also shows the smallest F0 difference between voiced and voiceless cases. However, as with F0 lowering, there seem to be language or speaker-based differences. Löfqvist et al's results are consistent with an electromyographic study by Dixit and MacNeilage (1980) of one Hindi speaker, but Collier, Lisker, Hirose and Ushijima (1979) report no relevant difference between the voiced and voiceless stops and fricatives of a speaker of Dutch.

There thus seem to be a number of mechanisms which can result in higher F0 after a voiceless consonant than after a voiced one. F0 lowering with voicing may be a consequence of a lowered larynx position, or due to the aerodynamic effects of a supraglottal constriction. Aerodynamic effects may raise F0 after voiceless consonants with an open glottis, and devoicing strategies that involve tensing the vocal folds will also be likely to raise F0 after voiceless consonants. The particular effects that come into play will produce somewhat differing patterns both with respect to the amount of F0 difference and the classes of sounds involved, but in no case will F0 be higher after a voiced consonant than after a voiceless one. The variability noted in the correlation of consonant voicing and lowered F0 is one factor that lead Kingston and Diehl (1994) to suggest that there is much more of a language-specific, learned, nature to this correlation, rather than a universal basis for it. (They interpret several other proposed phonetic universals in the same way).

3.3 Universals related to consonant place.

Several observational universals relate place of articulation and the duration of consonants or their environments. Most of the measurements related to this issue have been on stops; the longer closure of bilabial than of velar stops was mentioned in section 2 above. It has been informally suggested that this pattern is connected to the air pressure in the cavity behind the closure. The smaller this cavity is, the more rapidly the pressure will rise following closure and the sooner intraoral pressure will reach equality with pulmonic pressure (primarily because smaller cavities have smaller surface area). If the consonant gesture is timed in some way that directly relates to the time of the pressure peak, then broadly speaking, the further back in the oral cavity a stop closure is formed, the shorter its acoustic closure duration will be. An experiment by Ohala and Riordan (1979, see also Ohala 1983), showed that intraoral pressure becomes high enough to suppress voicing on average about 25 ms earlier for alveolar /d/ than for /b/, and about 15 ms earlier for /g/ than for /d/.

Comparing English bilabial, alveolar and velar stops in word-medial position of disyllabic nonsense words with second syllable stress, Stathopoulos and Weismer (1983) did find a monotonic relationship between the backness of the place and the shortness of the closure, as shown in Table 3. A similar pattern has been found, at least in certain environments, in other languages, for example in Standard Chinese word-initial unaspirated stops (Ren 1985), Swedish word-final short stops (Elert 1965: 143) and Florentine Italian intervocalic voiced stops following stress (Vagges, Ferrero, Magno-Caldognetto and Lavagnoli 1978). However, in many studies coronal stops have shorter closure durations than velars while labials remain longest. Byrd (1993) found this overall pattern in TIMIT, the large American English read-sentence database including 630 speakers, as also shown in Table 3 (note that the flap allophones of /t, d/ are excluded from this count).

Table 3. Stop consonant closure durations (in ms) and place of articulation.

English (6 speakers) (Stathopoulos and Weismer 1983)				**English (TIMIT: 630 speakers)** (Byrd 1993)			
p-	96	b-	92	p	69	b	64
t-	82	d-	76	t	53	d	52
k-	72	g-	68	k	60	g	54

This alternative pattern might be attributed to differences in the compressibility of the articulators. Bilabials are formed by closure between two soft surfaces which undergo a good deal of compression as they contact, velars involve contact between two somewhat compressible articulators — the tongue back and the soft palate — whereas dentals and alveolars involve a hard surface on one side of the contact. When the upper articulator is a yielding surface, there is likely to be a less rapid rebound from the ballistic movement that creates the contact, just as a ball bounces back more rapidly from a hard surface than a soft one. Air pressure and tissue compressibility effects may interact to create the pattern noted by Fischer-Jørgensen (1964) for Danish, in which bilabials are longest in their closure duration but alveolars and velars are of comparable duration. Since many factors affect durations, and there are also language-specific positional variations in stops to consider, it is often the case that inconsistent rankings are found even within a single language. Languages with multiple places for coronal stops, such as those of Australia, provide particular challenges for interpretation, as in the Rembarrnga data provided by McKay (1980) in which palato-alveolar stops are the longest, and postalveolars the shortest.

A more uniform observation is that the delay in voice onset following the release of a stop is dependent on how far back the place of articulation of the stop is; the further back, the longer the VOT. The English data in Byrd (1993) agree with this, with overall mean VOT for bilabials of 32 ms, for alveolars of 39 ms and for velars of 46 ms. Similar rankings are found in a wide range of languages, although results are not uniform. In general, voiceless bilabial plosives do have the shortest VOT, and velar VOT's are consistently longer than bilabials. However, uvulars are not consistently longer than velars, and the VOT of stops in the coronal region is highly variable, depending both on the part of the tongue used to form the constriction and its profile behind the constriction, as well as on what seem to be controlled differences in the way that the release itself is made (Ladefoged and Maddieson 1996). As noted in section 2, place-dependent VOT durations may have different causes under different circumstances. In the case of unaspirated stops, a short VOT for bilabials may be due to the rapid separation of the lips at release, both of them being mobile and moving away from one another, and accelerated by the effect of jaw opening. A long VOT for velars may result from the relatively great length of the contact in the sagittal plane. The long closure separates more slowly, often even reclosing momentarily to produce the 'double bursts' characteristic of velars. These different speeds of separation mean that the high intraoral pressure is dissipated more rapidly at the release of a bilabial stop than at the release of a velar. Hence, the transglottal pressure difference required for voicing to be (re-)initiated is reached earlier. However, as also noted above, explanations based on rate of separation do not apply well to aspirated stops, since the articulators are fully separated long before the aspiration ends.

3.4 Universals related to prosodic units.

A number of universal tendencies can be observed in relation to the prosodic structure of languages. A very widespread pattern is that some phonetic components are longer in a position before a boundary, such as at the end of a sentence or an intonational phrase, than they are in other positions. Beckman and Edwards (1990) are able to separate out phrase-final and word-final lengthening processes in English. Lengthening in these types of positions functions as a means of demarcating the ends of constituents, and presumably has a perceptual rather than a physiological basis. A longer syllable at the end of a constituent not only directly helps to mark the end of the constituent but also provides a better host for other markers of constituent structure such as tonal or voice quality changes, for example, the high-tone question-marker of Hausa (Lindau 1986). Pre-boundary lengthening may apply to a segment or syllable not in absolute final position, as in many Bantu languages which lengthen the penultimate syllable at the ends of sentences and, within sentences, at the ends of some major clauses. In this case, the lengthening may be positioned earlier in order to avoid the strongest impact of amplitude and pitch declination in the absolute final position.

Declination refers to an underlying downward trend of some trait over the length of an utterance or part of an utterance. Declination of F0 and overall acoustic amplitude appear to be universal, and there are some indications that declination of articulatory amplitude may also be typical (Vayra and Fowler 1992, Krakow 1993). Local perturbations – for example, for stressed syllables – may be superimposed on this overall decline, but successive local perturbations themselves often conform to a pattern of declination. In almost every language in which pitch patterns have been studied over utterances longer than word-length, the F0 pattern follows an overall downward trend, at least within certain spans and at least for some unmarked utterance type. It seems likely that this declination is, broadly speaking, associated with the baseline decline in the subglottal pressure that occurs as lung volume decreases over the course of an utterance. Important experiments reported in Gelfer, Harris, Collier and Baer (1985) and Gelfer, Harris and Baer (1987) showed that the Dutch-speaking subjects they studied controlled the rate of subglottal pressure decrement, as it remained constant with different rates of lung volume decrement arising from use of low air flow segments (/m/) versus high air flow segments (/f/). F0 declination was correlated with the declination of subglottal pressure, not with lung volume declination, nor with activity of the cricothyroid muscle. They therefore conclude that subglottal declination "is a controlled variable in sentence production, and that F0 declination is a consequence." The same authors conclude that the overall decline of acoustic amplitude is also a consequence of the declination of subglottal pressure. This effect, combined with anticipation of the open glottal position of normal respiration, produces a common tendency for elements in utterance-final position to become devoiced.

Phonetic universals have also been sought for smaller prosodic units such as the syllable, but few of the marks of syllable affiliation have been found to have cross-linguistic generality. One possible exception is shortening in closed syllables. In many languages a vowel in a closed syllable is shorter than a matched vowel in an open syllable (Maddieson 1985). Thus, the /ei/ in *'grey towel'* will tend to be longer than that in *'great owl'*. This effect may be observed most readily in languages where medial single and geminate consonants of the same type contrast (assuming that the geminate closes the preceding syllable). In this environment the influence of syllabification is isolated from other durational modifications. Some results for Italian, where this effect is very strong, are given in Table 4. Since there are languages in which geminate consonants do not have a shorter preceding vowel than single consonants, such as Japanese (Homma 1981, Smith 1995) and perhaps Sinhala (Letterman 1994), it is clear that this is not an automatic effect of the syllabification. As suggested above, what may be at work here is a strategy for keeping words with equal syllable count similar in overall duration (A possible advantage of doing this would be to keep the information density in the signal more uniform.)

Table 4. Duration of stressed /a/ before single and geminate consonants in Italian (Mean of three speakers, after Farnetani and Kori (1986))

Consonant type	**Vowel duration**	
	before single C	before geminate C
-l-	211	133
-n-	212	138
-s-	222	129

3.5 Some general comments

The various observational universals described briefly in the preceding sections illustrate a variety of situations. In some cases, such as F0 raising after voiceless consonants and place-dependant VOT differences of stops, a number of 'mechanical' factors work together to produce phonetic patterns that are broadly uniform in direction. In others, such as place-dependant patterns of stop closure duration, competing factors lead to expectation of conflicting trends. For some cases, such as final lengthening, a perceptual basis for a pattern seems more explanatory than an aerodynamic or physiological one. All these observations illustrate the point that 'universals' are not fixed attributes of languages. Rather, they are reflections of the overall context within which language is produced and heard. This is a fluctuating and dynamic set of conditions of which mechanical and perceptual constraints are only a part. Although these conditions set some boundaries and shape more likely and less likely outcomes, they do not result in linguistic homogeneity.

4. Ecological models

A somewhat different perspective on the issue of phonetic universals is based on what is here called an ecological view of the matter. Rather than considering primarily factors such as the physiological and aerodynamic conditions that rule during the production of a given sound or utterance, or the perceptual processing of a given sound or utterance, the ecological perspective considers a larger view of the functionality of language. Spoken language must, among other requirements, be constructed to conform to two overarching principles which might be labeled Contrastivity and Connectedness. A language must show variation in sound – rather than being an undifferentiated noise – if it is to convey any other message than simply "I'm speaking". It must also be possible to identify recurrent parts of the signal and recognize them as representing the same word (or morpheme or other meaningful constituent). That is, both speakers and listeners need to be able to identify the same message as being the same, and different messages as being different. These requirements create the property that is usually labeled phonological contrast; utterances must contain parts that are differentiated from each other and recognizable when they recur. The need for contrastivity is responsible for such things as the alternation of louder and quieter sounds that is the basis of syllabification and for the selection of elements with good characteristic 'signatures' in both their motor and auditory patterns so that they may be recognized and memorized.

Equally, a language needs to be produced as a continuous stream, its parts connected to each other just as essentially as they must be differentiated from each other. Moreover, since the position in which a given word or other element will occur is variable, the form of any item must be adapted for variable environments. These considerations place some limits on the degree to which articulatory displacements are made, especially in adjacent parts of an utterance. Moving an articulator to an extreme displacement (e.g. sticking the tongue out as far as possible between the teeth) makes it harder to connect to the next position – unless it is the same, and variability of context ensures that it will more frequently not be the same. Connectedness therefore favors moderate articulatory displacements over extreme ones. However, some variation in articulatory instantiation of a given word facilitates its concatenation with other items in contexts that themselves vary. Nonetheless, beyond a certain level, variation begins to conflict with the need to be able to identify repetitions of the same item. Variation will therefore tend to remain within limited bounds.

Note that, although these comments could be made about a segmental level of analysis, and will be discussed below in the light of such an analysis, it is not necessary to presuppose that any such level exists. The observations above hold whether one regards segments as 'emergents' that arise from a self-organizing system (Lindblom 1992), whether one believes that the human system is pre-adapted in some way to a segmental analysis of speech (Ohala 1992), or whether one believes that there is no level of segmental organization (Local 1992). Connectedness requires flexibility in production, and favors less extreme articulations under any interpretation of the units involved.

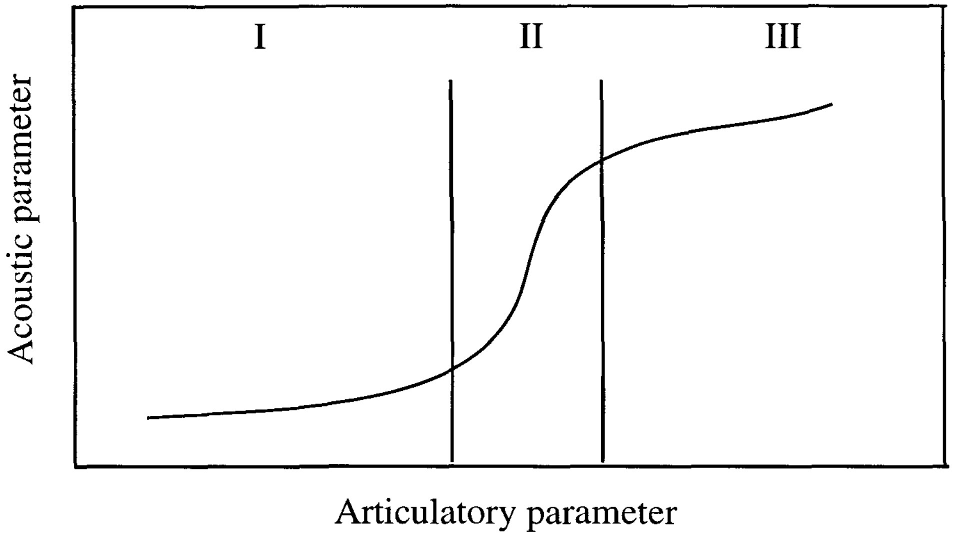


Figure 4. Schematic relationship between an articulatory and an acoustic parameter as discussed by Stevens (1972, 1989).

Contrastivity and connectedness as described here are related to the idea that, in their phonetic structure, languages must maintain a balance between auditory distinctiveness and ease of articulation. Lindblom (1986, 1990) has argued that the interaction of these two principles generates many of the patterns described as markedness relations between different segments or segment inventories. We argue that there is a requirement for articulatory as well as auditory distinctiveness, and suggest a different perspective on the question of ease of articulation, in which the value of 'unity in diversity' is stressed. The Quantal Theory of Stevens (1972, 1989) provides one window on the issue of contrastivity. This proposes that nonlinearities in the relationships between articulatory parameters and the resulting acoustic/auditory responses are a key to this issue. In figure 4, based on Stevens (1989: 4), the horizontal axis represents changes in some articulatory parameter on a linear scale, for example, steps in increasing the distance between the lips. The vertical scale represents changes in some acoustic or auditory parameter, for example, the overall amplitude of the signal, in response to this change. Region II is an area where there are large changes in the acoustics for small shifts in articulation, as at the release of a labial closure. Within regions I and III there is relatively little difference in the acoustics for articulatory shifts of equivalent size, but the difference between Region I and Region III is large. In the example we have used this might correspond to the acoustic difference between stop and vowel. Over a small but crucial range of lip movement the acoustic output demonstrates a relatively sharp change in amplitude. Many other pairs of parameters are similarly related. Stevens suggests the general principle that segment sequences are selected so that they cross regions such as Region II in this figure, producing rapid changes that serve as landmarks in the acoustic stream.

In whatever way concepts such as connectedness and contrastivity are formulated, they are attempts to capture some of the ambient factors that govern the overall design of the sound patterns of languages. Such factors in the 'ecology' of language shape phoneme frequency, the structure of phoneme inventories, and the pattern of contrast distribution, among other matters. For example it has often been noted that phoneme inventories are constructed in such a way as to appear built on a foundation of use of basic parameters, which are found in almost all languages including those with very small numbers of segmental distinctions. As increasing numbers of distinctions are made, additional ways of making them are added. For example, we may observe that vowel systems start from height variations; a few languages have only two or three distinctive vowels, and contrast these on the height dimension alone. No language is known which does not have some distinctions of height. Other languages with as few as three vowels and all those known with five or more also show front/back distinctions among their vowels, with typically a redundant association between rounding and backness which enhances the acoustic distinction (contrastivity) between front and back. Only languages with relatively large numbers of distinct vowel qualities disrupt this association between rounding and backness and permit these parameters to be independently varied, yielding front rounded and back unrounded vowels in contrast with the more frequently encountered front unrounded and back rounded types. Nasalization, pharyngeal volume, and voice quality differences are other more elaborated distinctions among vowels (Ladefoged and Maddieson 1996).

There is generally a good correlation between the relative frequency of a segment type in terms of its appearance in the inventories of languages around the world and its frequency within the lexicon of particular languages. The UPSID archive (Maddieson 1984a), currently containing data on 453 languages, provides a basis for determing which are the most frequent sound types in the world's languages. It may be inferred that these sound types meet the ecological criteria better than others (though their frequency is not by itself evidence of this fact). Among consonants, 80% or more of languages have p, *t, k, m, *n, *s, j (where * indicates a class of coronal consonants that includes dental and alveolar places). Over 50% have b, *d, ɡ, ŋ, h, *l, w, and some form of *r. In a count of lexical items in a sample of 25 languages of widely varied genetic and geographical groupings the most common individual syllable-onset consonants were /k/ (10 languages), /*t/ (5 languages), /l/, /s/ (2 languages each) and /m/, /n/, /d/, /h/, /r/ and /tɕ/ (1 language each). All except /tɕ/ are among the most common consonants cross-linguistically, but an interesting fact that emerges from this study is the greater favoring of /k/ relative to other common consonants. In a similar way, although /i/, /a/ and /u/ are all found in over 80% of the UPSID language sample, lexical counts show that, in many languages, /a/ is far more frequent than any other vowel (Maddieson 1992). We might therefore say that the optimal syllable is /ka/.

The listing of most common consonants also illustrates a basic structural principle; inventories tend to be built by the intersection of repeated characteristics. For example, nine of the most commmon consonants can be laid out on a grid with three rows, as in Table 5. By replicating essentially identical articulatory gestures, the task of learning motor 'images' of sounds is reduced in complexity, and we may also infer that the places of articulation represented are those that are most easily connected to other sounds in the stream of speech. Note that there is one 'modal' place represented for each of the three main independent oral articulators, the lips, tongue tip and tongue body (granted the conflation of dental and alveolar). More extreme coronal or dorsal articulations, such as interdental, linguo-labial, or uvular are rare by comparison. This again illustrates that the construction of inventories starts from basic elements and extends to include less basic elements as inventory size increases (Lindblom and Maddieson 1988).

Table 5. A grid of some of the most common consonants.

	bilabial	dental/alveolar	velar
voiceless plosive	**p**	**t**	**k**
voiced plosive	**b**	**d**	**g**
nasal	**m**	**n**	**ŋ**

A related principle that that has been widely discussed is that of ‘dispersion’. In early formulations (Liljencrants and Lindblom 1972, Lindblom 1986), the idea was that contrastive elements within a given phonetic domain should be expected to be *maximally* separated from each other. For example, a three vowel system would be expected to consist of the vowels /i, a, u/, where /i, u/ have the lowest possible first formant values but differ by having opposite extreme values of the second formant, and /a/ has the highest possible first formant. If a language had a larger number of distinct vowels, /i, a, u/ would remain in the same position and the others would be added between the extreme values, giving, say, a system containing /i, e, ɛ, a, ɔ, o, u/. Analysis of tone systems with different numbers of level tones suggests instead that contrastive elements are not maximally separated, and that increasing numbers of contrasts may be accommodated by moving the end-points of the scale further apart rather than by fitting more steps in between fixed end-points. Data in Table 6 from two Hausa-Nupe bilinguals (hence, controlled for speaker differences in pitch range) shows that the two lower tones of Nupe are essentially on the same levels as the two tones of Hausa, but the third tone of Nupe — the high tone— is substantially higher.

Table 6. Mean F0 (in Hz) of tones in matched environments in Hausa and Nupe as spoken by two bilingual speakers (after Maddieson 1991).

Hausa			**Nupe**		
			high	135	} 14
high	123	} 20	**mid**	121	} 23
low	103		**low**	98	

That something similar is found with vowels is indicated by Figure 5, comparing the values of the first two formants of /i, a, u/ in a dialect of Bavarian German with many other vowels and in Tausug, an Austronesian langugage of the Philippines, which has only these three vowels. Although the vowels being compared have been written the same in the two languages, the high vowels are markedly lower in Tausug than in the German dialect. Although few such comparisons have been made, it seems likely that when there are few vowels the phonetic realizations of the most distinct vowels are in general closer together in a space defined by the first two formants than is the case in a language with more vowels. These two examples show that contrastivity is satisfied by less than a maximal separation; however, the difference between the end-points of any scale looked at cannot be extended indefinitely far apart. Comfortable limits are soon reached (with three or four tone levels and five or six vowels), and contrast will then tend to be extended by use of other parameters (duration, dynamic patterns, differences in source spectrum, etc).

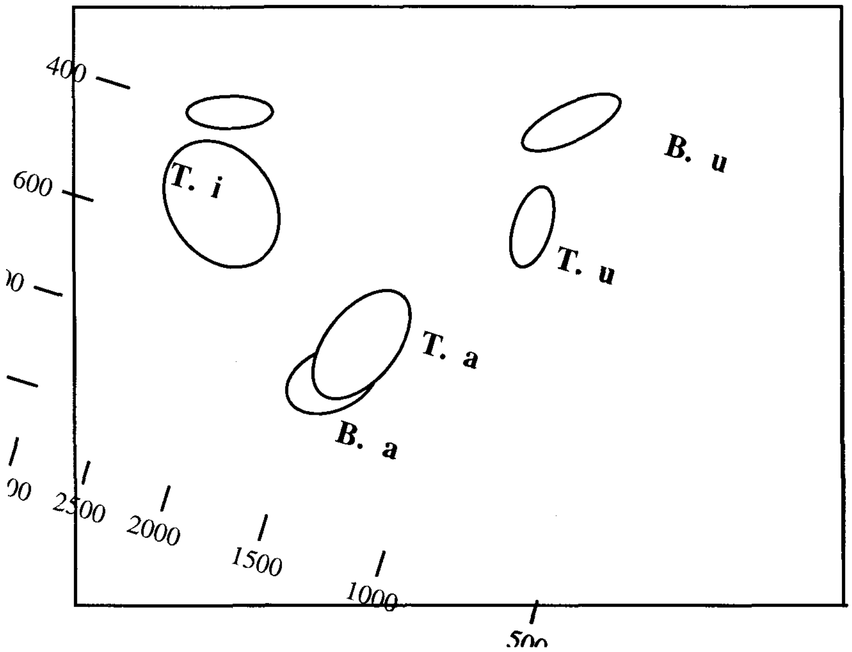


Figure 5. The three most extreme vowels of a Bavarian German dialect with 13 distinct long vowels (B.) and the three vowels of Tausug (T.) plotted in F1/F2 space (based on Disner 1983). The ellipses enclose approximately 95% of the variance in the data, which is from 8 male speakers of Bavarian and 4 of Tausug.

This sections has outlined some of the major ecological considerations that the sound systems of languages must meet and briefly discussed a few of the consequences of these requirements in terms of observed patterns across languages. Research on 'ecological' universals is perhaps less developed than research on 'mechanistic' universals. This may be because the theories that describe the ecological setting are broader in scope than those that address 'mechanistic' universals, and therefore developing tests of their predictive power is more challenging.

References

Beckman, Mary E. and Jan Edwards. 1990. "Lengthenings and shortenings and the nature of prosodic constituency." In *Papers in Laboratory Phonology 1*, ed. by J. Kingston and M. E. Beckman. Cambridge University Press, Cambridge: 152-178.

Bickley, Corinne B. and Stevens, K. N. 1986. "Effects of a vocal-tract constriction on the glottal source: experimental and modelling studies." *Journal of Phonetics* 14: 373-382.

Bickley, Corinne B. and Stevens, K. N. 1991. "Effects of a vocal-tract constriction on the glottal source: data from voiced consonants." In *Vocal Fold Physiology: Laryngeal Function in Phonation and Respiration*, ed. by K. S. Harrris, T. Baer and C. Sasaki. College Hill Press, San Diego: 239-253.

Byrd, Dani. 1993. 54,000 stops. *UCLA Working Papers in Phonetics* 83: 97-115.

Catford, J. C. 1977. *Fundamental Problems in Phonetics.* Indiana University Press, Bloomington, IN.

Chen, Matthew. 1970. "Vowel length variation as a function of the voicing of the consonant environment." *Phonetica* 22: 129-159.

Collier, René, Leigh Lisker, Hajime Hirose and Tatsujiro Ushujima. 1979. "Voicing in intervocalic stops and fricatives in Dutch." *Journal of Phonetics* 7: 357-373.

Disner, Sandra F. 1983. Vowel Quality: The Relation Between Universal and Language-Specific Factors. *UCLA Working Papers in Phonetics* 58.

Dunn, Margaret. 1993. *The Phonetics and Phonology of Geminate Consonants: A Production Study*. Unpublished Ph. D. dissertation, Yale University, New Haven.

Elert, Claes-Christian. 1964. *Phonologic Studies of Quantity in Swedish.* Almqvist and Wiksell, Stockholm.

Ewan, William G. and Robert Krones. 1974. Measuring larynx movement using the thyroumbrometer. *Journal of Phonetics* 2: 327-335.

Farnetani, Edda and Kori, Shiro. 1986. Effects of syllable and word structure on segmental durations in spoken Italian. *Speech Communication* 5: 17-34.

Fischer-Jørgensen, Eli. 1964. "Sound duration and place of articulation in Danish." *Zeitschrift fur Sprachwissenschaft und Kommunikationsforschung* 17: 175-207.

Fowler, Carol A. 1992. Vowel duration and closure duration in voiced and unvoiced stops: there are no contrast effects here. *Journal of Phonetics* 20: 143-165.

Gandour, Jack and Ian Maddieson. 1976. "Measuring larynx height in standard Thai using the cricothyrometer." *Phonetica* 33: 241-267

Gelfer, Carole E., Harris, Katherine S., Collier, René and Baer, Thomas. 1985. Is declination actively controlled? In *Vocal fold physiology: Physiology and Biophysics of the Voice*, ed by I. Titze. Iowa University Press, Iowa City: 422-435.

Gelfer, Carole E., Harris, Katherine S., and Baer, Thomas. 1987. Controlled variables in sentence intonation. In *Vocal Fold Physiology: Laryngeal function in Phonation and Respiration*, ed by T. Baer, C. Sasaki and K. S. Harris. College Hill Press, Boston: xxx-xxx.

Hoemeke, Kathryn A. and Diehl, Randy L. 1994. Perception of vowel height: the role of F_1-F_0 distance. *Journal of the Acoustical Society of America* 96: 661-674.

Hombert, Jean-Marie. 1978. "Consonant types, vowel quality, and tone." In *Tone: A Linguistic Survey*, ed by V. Fromkin. Academic Press, New York: 77-111.

Homma, Yayoi. 1981. "Durational relationships between Japanese stops and vowels." *Journal of Phonetics* 9: 273-281.

Honda, Kiyoshi and Fujimura, Osamu. 1991. "Intrinsic vowel F_0 and phrase-final lowering: phonological versus biological explanations." In Gauffin, J. and Hammarberg, B. *Vocal Fold Physiology: Acoustic, Perceptual and Physiological Aspects of Voice Mechanisms*. Singular Press, San Diego: 149-157.

Jaeger, Jeri J. 1978. "Speech aerodynamics and phonological universals." *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society* : 311-329.

Javkin, Hector R. 1976. "The perceptual basis of vowel duration differences associated with the voiced/voiceless distinction." *Report of the Phonology Laboratory* (University of California, Berkeley) 1: 78-92

Kingston, John, and Diehl, R. L. 1994. "Phonetic knowledge." *Language* 70: 419-454.

Kluender, Keith R., Diehl, R. L. and Wright, B. A. 1988. "Vowel-length differences before voiced and voiceless consonants: an auditory explanation." *Journal of Phonetics* 16: 153-169.

Krakow, Rena A. 1993. "Nonsegmental influences on velum movement patterns: syllables, sentence, stress, and speaking rate." In *Nasals, Nasalization, and the Velum*, ed by M. K. Huffman, and R. A. Krakow. Academic Press, San Diego: 87-116.

Kuehn, D. P. and Moll, K. L. 1976. "A cineradiographic study of VC and CV articulatory velocities." *Journal of Phonetics* 4, 303-320. Reprinted in Kent, R. D., Atal, B. S. and Miller, J. L. 1991. *Papers in Speech Communication: Speech Production.* Acoustical Society of America, Woodbury, NY: 527-565.

Ladefoged, Peter. 1967. *Three Areas of Experimental Phonetics.* Oxford University Press, London.

Ladefoged, Peter and Ian Maddieson. 1996. *Sounds of the World's Languages.* Blackwells, Oxford

Letterman, Rebecca. 1994. "A phonetic study of Sinhala syllable rhymes." *Working Papers of the Cornell Phonetics Laboratory* 9: 155-181.

Liljencrants, Johan and Björn Lindblom. 1972. "Numerical simulation of vowel quality systems: The role of perceptual contrast." *Language* 48: 839-862.

Lindau, Mona. 1986. "Testing a model of intonation in a tone language." *Journal of the Acoustical Society of America* 80: 757-764.

Lindblom. Björn. 1967. "Vowel duration and a model of lip-mandible coordination." *Speech Transmission Laboratory Quarterly Progress and Status Report (Royal Institute of Technology, Stockholm)* : 1-29.

Lindblom, Björn. 1986. "Phonetic universals in vowel systems." In *Experimental Phonology,* ed by J. J. Ohala and J. Jaeger. Academic Press, Orlando: 13-44.

Lindblom, Björn. 1990. "Models of phonetic variation and selection." *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PERILUS)* 11: 65-100.

Lindblom, Björn. 1992. "Phonological units as adaptive emergents of lexical development." In *Phonological Development: Models, Research, Implications*, ed by C. F. Ferguson, L. Menn and C. Stoel-Gammon. York Press, Timonium MD: 131-164.

Lindblom, B. and I. Maddieson. 1988. "Phonetic universals in consonant systems." In *Language, Speech and Mind: Studies in Honor of Victoria A. Fromkin*, ed by L. M. Hyman and C. N. Li. Routledge, London and New York: 62-80.

Lisker, Leigh and Abramson, A. S. 1964. "A cross-language study of voicing in initial stops: Acoustical measurements." *Word* 20: 384-422. Reprinted in Kent, R. D., Atal, B. S. and Miller, J. L. 1991. *Papers in Speech Communication: Speech Production.* Acoustical Society of America, Woodbury, NY: 671-688.

Local, John. 1992. "Modeling assimilation in non-segmental, rule-free synthesis." In *Papers in Laboratory Phonology 2: Gesture, Segment, Prosody*, ed. by G.J. Docherty and D. R. Ladd. Cambridge University Press, Cambridge :190-223.

Löfqvist, Anders, Thomas Baer, Nancy S. McGarr and Robin Seider Story. 1989. "The cricothyroid muscle in voicing control." *Journal of the Acoustical Society of America* 85: 1314-1321.

McKay, G. R. 1980. "Medial gemination in Rembarrnga: A spectrographic study." *J.ournal of Phonetics* 8: 343-352.

Maddieson, Ian. 1984a. *Patterns of Sounds*. Cambridge University Press, Cambridge.

Maddieson, Ian. 1984b. "The effects on Fo of a voicing distinction in sonorants and their implications for a theory of tonogenesis." *Journal of Phonetics* 12: 9-15.

Maddieson, Ian. 1985. "Phonetic cues to syllabification." In *Phonetic Linguistics: Essays in Honor of Peter Ladefoged,* ed by V. A. Fromkin. Academic Press, New York: 203-221.

Maddieson, Ian. 1991. "Tone spacing." In *Festschrift for Professor Jack Carnochan* (ed. J. Kelly and J. Local). *York Papers in Linguistics* 15 University of York, Heslington. 149-175.

Maddieson, Ian. 1992. "The structure of segment sequences." In *Proceedings of the 1992 International Conference on Spoken Language Processing*, ed by J. J. Ohala et al. University of Alberta, Banff, Alberta: Addendum 1-4.

Mohr, Burkhardt. 1971. "Intrinsic variations in the speech signal." *Phonetica* 23: 65-93.

Ohala, John J. 1978. "The production of tone." In Fromkin, V. A. *Tone: A Linguistic Survey.* Academic Press, New York: 5-39.

Ohala, John J. 1983. "The origin of sound patterns in vocal tract constraints." In *The Production of Speech*, ed by P. F. MacNeilage. Springer-Verlag, New York: 189-216.

Ohala, John J. 1992. "The segment: primitive or derived?" In *Papers in Laboratory Phonology 2: Gesture, Segment, Prosody*, ed by G. J. Docherty and D. R. Ladd. Cambridge University Press, Cambridge: 166-183.

Ohala, John J. and Carol Riordan. 1979. "Passive vocal tract enlargement during voiced stops." In *Speech Communication Papers*, ed by J. J. Wolf and D. H. Klatt. Acoustical Society of America, New York: 89-92.

Ohala, Manjari. 1980. *Aspects of Hindi Phonology*. Motilal Banarsidass, New Delhi.

Raphael, Lawrence J., Dorman, M. F., Freeman, F. and Tobin , C. 1975. "Vowel and nasal duration as cues to voicing in word-final stop consonants: Spectrographic and perceptual studies." *Journal of Speech and Hearing Research* 18: 389-400.

Ren, Hongmo. 1985. "Linguistically conditioned duration rules in a timing model for Chinese." *UCLA Working Papers in Phonetics* 62: 34-5o.

Riordan, Carol J. 1980. "Larynx height during English stop consonants." *Journal of Phonetics* 8: 353-360.

Smith, Caroline L. 1995. "Prosodic patterns in the coordination of vowel and consonant gestures." *In Phonology and Phonetic Evidence: Papers in Laboratory Phonology* 4, ed by B. A. Connell and A. Arvaniti. Cambridge University Press, Cambridge: 205-222.

Sonesson, B. 1982. "Vocal fold kinesiology." In *Speech Motor Control,* ed by S. Grillner, B. Lindblom, J. Lubker and A. Persson. Pergamon Press, Oxford: 113-117.

Stathopoulos, Elaine T. and Gary Weismer. 1983. "Closure duration of stop consonants." *Journal of Phonetics* 11: 395-400.

Stevens, Kenneth N. 1972. "The quantal nature of speech: Evidence from articulatory-acoustic data." In *Human communication: A unified view*, ed by E. E. David and P. B. Denes. McGraw-Hill, New York: 51-66.

Stevens, Kenneth N. 1989. "On the quantal nature of speech." *Journal of Phonetics* 17: 3-46.

Umeda, Noriko. 1977. "Consonant duration in American English." *Journal of the Acoustical Society of America* 61: 846-858.

Vagges, K., Ferrero, F. E., Magno-Caldognetto, E., and Lavagnoli, C.. 1978. Some acoustic characteristics of Italian consonants. *Journal of Italian Linguistics* 3: 68-85.

Vayra, Mario, and Fowler, Carol A. 1992. "Declination of supralaryngeal gestures in spoken Italian." *Phonetica* 49: 48-60.

Weismer, G. 1980. "Control of the voicing distinction for intervocalic stops and fricatives: Some data and theoretical considerations." *Journal of Phonetics,* 8, 427-438.

Whalen, Douglas and Levitt, Andrea. 1995. "The intrinsic pitch of vowels." *Journal of Phonetics* 24: 349-366.