THE PHONETIC CONTRAST OF KOREAN OBSTRUENTS

Jonathan D. Wright

A DISSERTATION

in

Linguistics

Presented to the Faculties of the University of Pennsylvania in Partial

Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2007

Mark Liberman, Supervisor of Dissertation

Eugene Buckley, Graduate Group Chair

Acknowledgements

ABSTRACT

THE PHONETIC CONTRAST OF KOREAN OBSTRUENTS

Jonathan D. Wright

Supervisor: Mark Liberman

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Overview

1.1 Introduction

Korean is somewhat famous (among linguists) for its obstruents. It has a manner of articulation that is apparently unique, which provides for a three-way voiceless contrast, also apparently unique (if languages with ejectives are excluded). What's less well-known is that Korean obstruents have a direct influence on phrasal pitch. While consonant-tone interaction in general is not unique, the specific interaction found in Korean may be. Finally, young speakers from Seoul seem to be participating in a change in progress regarding the voiceless stops. This segmental change in progress is bringing the tonal contrast to the foreground, raising questions about tonogenesis and the phonological status of Korean tone.

The broad aim of this dissertation is to bring together these threads into a coherent picture, and shed further light on that picture through phonetic analysis and experimentation. The remainder of this chapter is a summary of that picture, drawing from both my work and the literature, and in many ways is conclusory as well as introductory. Not only will this chapter provide a foundation for the other chapters, but readers not interested in the details will get a fairly complete picture of this topic from this chapter alone. Chapter 2 is a systematic literature review of relevant research, and the following chapters describe my own phonetic experiments. The final chapter provides more conclusions and future directions. This disseration is foundational, too broad in scope to complete many of the threads of research that I began. Therefore future projects are sure to follow on the heels of this one.

1.2 Overview of the Relevant Aspects of Korean

1.2.1 Segmental Overview

The following table shows the Korean obstruents.

	plain	aspirated	tense
labial stop	ㅂ p	ㅍ p^h	ㅃ p*
velar stop	ㄱ k	ㅋ k^h	ㄲ k*
coronal stop	ㄷ t	ㅌ t^h	ㄸ t*
coronal affricate	ㅈ č	ㅊ $č^h$	ㅉ č*
coronal fricative		ㅅ s^h	ㅆ s*

There is no voicing contrast among the obstruents, although the plain series is subject to a voicing rule. Therefore Korean has an unusual three-way voiceless contrast with no ejectives and no contrastively voiced stops, and this is due to the apparently unique *tense* manner of articulation. The three manners of articulation have been described variously as follows:

1. plain: plain, lax, lenis, unaspirated, slightly aspirated, unmarked, underspecified for laryngeal features
2. aspirated: aspirated, heavily aspirated, tense, stiff glottis, spread glottis
3. tense: tense, forced, reinforced, fortis, stiff glottis, constricted glottis, glottalized, geminate

Aside from descriptive uses, authors tend to pick one term for each stop simply to use as labels, for example *fortis, lenis, and aspirated*, or *tense, lax, and aspirated*, or *tense, slighly aspirated, and heavily aspirated*, etc. Since authors make different choices, terminology can at times become confusing. As you can see, I have chosen *tense, plain, and aspirated*; this choice was mainly based on minimizing confusion, and I have no fully logical rationale for the choice. I will attempt to stick to these labels as closely as possible.

The following word-level phonological rules apply in Korean[1].

1. Plain obstruents become voiced inter-sonorantly.
2. Plain obstruents become tense when preceded by an obstruent.
3. Plain stops become aspirated when adjacent to /h/.
4. In coda position, all osbtruents are neutralized to a homorganic voiceless unreleased stop.

This description of Korean obstruents is the standard, traditional description, capturing the surface alternations[2]. The correct underlying representations is a matter of much debate, as described in Chapter 2. However, this description will be the standard reference point for

[1] In addition, before /i/ coronal stops become affricates and /s/ moves from alveolar to alveo-palatal.
[2] Necessary rule or constraint orderings is beyond the scope of this thesis

all discussion herein. I call it the Received Representation for Korean obstruents. Crucial to this Received Representation is the description of these rules as word-level. We will see in the next section why word-level is not accurate, but in this section I will continue to use that description.

One interesting feature of Korean stops, and a central concern of this dissertation, is the apparent change in progress of aspiration in the Seoul dialect: for young Seoul speakers word-initial plain stops are often as heavily aspirated as the aspirated stops. Silva (2006) is probably the first definitive report of this change in progress[3]; his results and my own show that some speakers no longer have contrastive aspiration in initial position. Chapter 3 reports on the results of my experiment designed to examine this change in progress. Despite this change, for speakers that have no aspiration contrast, it's unclear whether a true positional neutralization has taken place. First, a closure duration contrast was observed in my experiment, aspirated stops having a longer closure duration, even when there was no aspiration contrast, for all but one speaker. Second, as many authors have observed, there is a tonal contrast; I will discuss tone below (p. 5).

Tense obstruents seem to be unique to Korean. Acoustically the tense stops are voiceless with a very small VOT, roughly 10 ms, and essentially an instantaneous rise to full amplitude of phonation. Articulatorily, the glottis is adducted and tensed immediately before stop release[4] (Kagaya, 1974). One question that may come to mind is how do they differ from ejectives, in terms of articulation. In other words, if the glottis is adducted, are we sure that these aren't just short ejectives? Kagaya (1974) notes that the glottis is not completely adducted and has a small gap, similar to the adducted glottis during phonation. This gap certainly allows airflow during /s*/ and /ch*/, and we can surmise that this glottal

[3]Silva (1992) is probably the first source that suggests such a change. Silva (2006) appears to be the first description of a trend in apparent time based on the author's own phonetic measurements of a large sample size.

[4]But not throughout the entire closure; see the discussion of Kagaya (1974) in Chapter 2.

configuration is what prevents a drift towards ejective articulation: it is qualitatively different. Tense stops also have relatively long closures, more akin to aspirated stops than plain stops in this respect.

The preceding descriptions specifically apply to word-initial position. Recall the Received Representation where tense and aspirated stops are unchanged word-internally while plain stops undergo alternations. But the situation is not that simple. Tense stops are notably longer word-internally suggesting to some authors that they are in fact geminates[5]. Aspirated stops are shorter word-internally, both in terms of closure duration and aspiration duration. In my opinion, this presents a peculiar problem for the overall analysis, since it can be difficult to separate allophonic alternations from the effects of prosody and speech rate. For example, aspirated stops have less aspiration word-internally, which is probably due to prosodic weakening, but could be analyzed as an unaspirated or weakly aspirated allophone. Recall that the plain stops are well-known to be (at least) slightly aspirated, so it seems that an allophonic change from aspirated to plain word-internally is fair game. In other words, the shorter aspiration *and* closure of word-internal aspirated stops makes them quite similar to word-initial plain stops. Plain stops tend to voice word-internally when between sonorants, but this voicing is not always complete and a positive VOT is sometimes seen. Particularly in elicited speech this can be seen, but this suggests it is a speech rate effect. A sufficiently long stop closure will cause the cessation of phonation due to supraglottal pressure, and doesn't mean the stop wasn't voiced phonologically. Therefore the analysis becomes murky depending on how one integrates speech rate and utterance types. For example, if the word-internal variants of aspirated stops are simply due to prosodic weakening, then perhaps plain stop voicing is also phonetic weakening and not allophonic variation.

[5]Note that orthographically they *are* geminates. I don't believe this is evidence for an analysis, but is perhaps relevant to the existence of geminate analyses to begin with.

One element not present in the Received Representation is tone. Even relatively old linguistic publications, say pre-1970, noted the following obstruent-pitch correlation: initial tense and aspirated obstruents have a higher associated pitch than plain obstruents. The nature of this obstruent-pitch correlation is not always clear in the literature, some authors attributing it to microprosody, to a local phonetic effect of phonation types, and others granting it phonological status. However, it must certainly be the latter. While the tone does vary by phonation type, it has a substantial effect on the following vowel and phrase that on the surface is as distinctive as a true lexical tone. In other words, the effect is clearly non-local (Jun, 1996b, p.46). Word-initial tense and aspirated obstruents, as well as /h/, induce a high tone over the following vowel. This pattern is true for Seoul Korean as well as other dialects (see **??**). See the following section on prosody for more discussion.

Plain stops and sonorants pattern together in terms of tone, correlating with a low relative to the high of tense obstruents, aspirated obstruents, and /h/. I will refer to these two groups of segments as L-segments and H-segments. By *segment* I am referring to the traditional melodic segments, not to any sort of autosegmental representation; the terms are descriptive, used to aid exposition. In a later chapter I demonstrate that the L-segments are homogeneous in terms of tone. Whether H-segments are truly homogenous is yet to be shown with certainty; Choi (2002) claims there is a difference for Seoul speakers between tense and aspirated segments.

1.2.2 Prosodic Properties

A discussion of Korean phrasal phonology should begin with Jun (1996b, 1998). Jun develops the concept of the Accentual Phrase (AP) for Korean[6], the domain for the assignment of tonal contours, analogous to the Japanese AP. The canonical form of the Seoul Korean AP, in Jun's terms, is a LHLH tonal contour. On the other hand, the canonical form for

[6]But she is not the first to suggest it.

APs that begin with H-segments is a HHLH contour. I say "canonical" because short APs have abbreviated contours. Analogous to the terms L-segment and H-segment, I will use the terms L-AP and H-AP to refer to these two basic AP types.

The Korean AP is between the Intonational Phrase and Phonological Word in Jun's theory, explictly standing in for the Phonological Phrase in the typical Prosodic Hierarchy (Selkirk, 1984). Jun's claim, which I support, is that the AP is in fact the domain for the segmental alternations described previously (p. 3). She chooses the term AP over Phonological Phrase because she is defining the domain based on tone[7] (Jun, 1996b, p.65). Therefore, what I have so far described as word-initial and word-internal will from now on be called AP-initial and AP-internal. Word-initial and word-internal will specifically refer to morpho-syntactic positions, not phonological ones.

As I said above, LHLH is the default tonal contour for Seoul Korean[8], but is only realized as such if the AP is long enough. Short APs have an abbreviated contour, roughly just LH[9]. My experience with laboratory data is that words of four or more syllables clearly show the full contour. The first LH and the second LH seem to anchor to the beginning and ending of the word, with long words having a gradual fall from H to L in the center, which is just how Jun (1996b, p. 56) describes the alignment. Furthermore, in laboratory data, there is a close correspondence between words and APs, in other words, a one-to-one correspondence for the most part. Figure 1.1 shows two APs representative of this, the former being an L-AP, the latter an H-AP. They are segmentally identical, except for the /p/ and /p^h/ that occur in initial position.

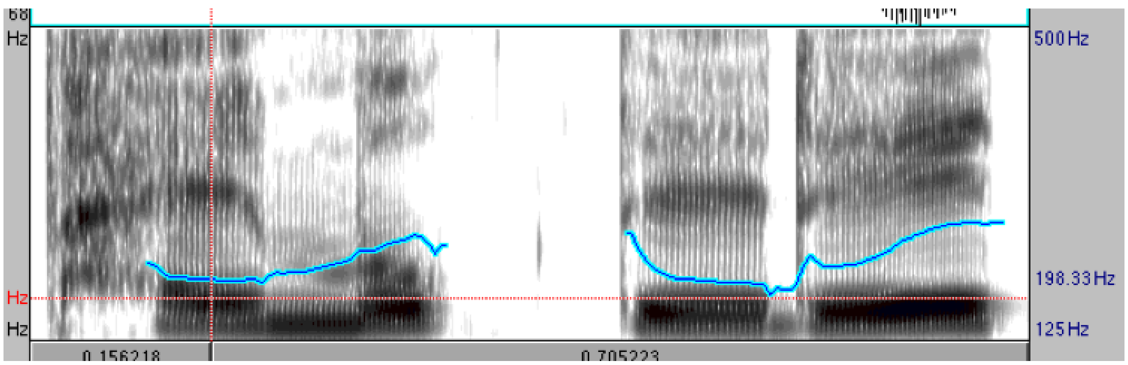
[7] I won't address issues of the syntax-phonology interface or prosodic hierarchy here, since they are tangential. The central claim of interest is that the same domain is used for tonal assignment and segmental alternations.

[8] In Chonnam Korean, the tone bearing unit is the mora rather than the syllable, and the basic tonal contours are LHL and HHL.

[9] Jun (1998) gives more details than that, and I refer the interested reader there. It is my current opinion that such distinctions as LLH and LHH are difficult to make in, for example, three syllable words that superficially have a LMH contour, and I haven't dwelled on such distinctions.

Figure 1.1: Canonical APs

발목들을 pal-mok-til-il



팔목들을 p^{h}al-mok-til-il

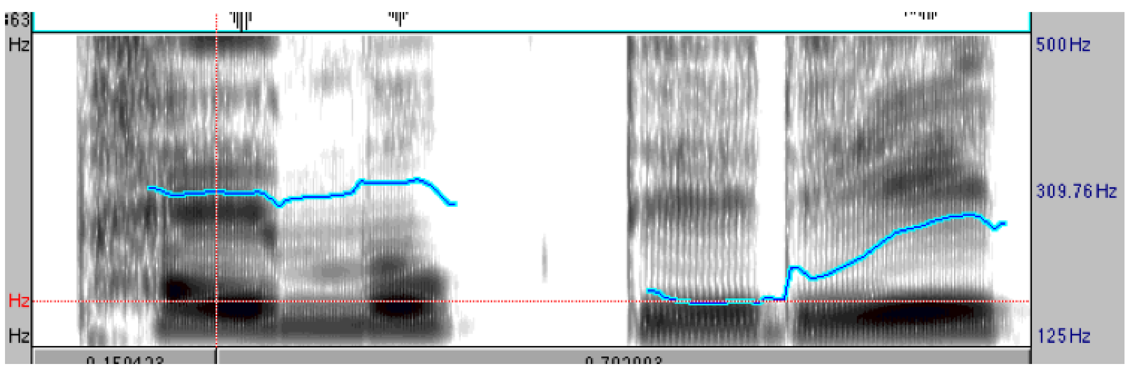
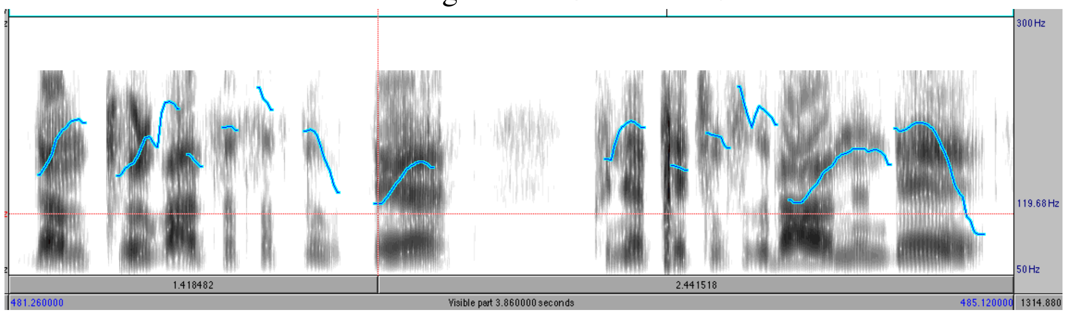


Figure 1.2: Utterance 481



Note that in both of these figures, there is a derived [t*], since there is the underlying sequence /kt/. The derived [t*] has no phonological impact on the tonal contour (although microprosodic effects are evident, see p. **??** for discussion on microprosody). As noted before, H-segments internal to the AP, whether underlying or derived, have no effect on the tonal contour. This conclusion is supported by the analysis of spontaneous speech in Chapter **??**. Figure 1.2 shows an underlying /t*/, marked by the red line, from the word ki-ttae.

Spontaneous speech clearly allows for multi-word APs, as Jun describes. However, her algorithm for tonal alignment is not allows obeyed, and the second position H of APs sometimes lies on the third syllable. This could simply reflect a speech rate effect, or it could indicate that tonal placement isn't based on syllable alignment. I don't offer an analysis on this point. But to reiterate another, more important, point: the morphological word is not, it seems, a relevant domain for the most salient aspects of Korean phonology. Rather, the AP is the domain for tonal assignment as well as the domain for segmental alternations.

Another important aspect of Jun's theory is that the segment-induced tone is phonological rather than purely phonetic, and she seems to be the first to say so. It's been established that different phonation types have different microprosodic effects on pitch (Hombert et al., 1979; Cristo and Hirst, 1986; Silverman, 1986; Kingston and Diehl, 1994). However, the tonal difference between L-APs and H-APs is not local to the initial consonants, and Jun (1996b, p.46) makes this clear.

There is an odd resistance in the literature to seeing segment-induced tone as a phonological effect, rather than a microprosodic one. I believe there are several reasons for this. First of all, it's just not widely known that this pattern exists, even among linguists. Japanese and Chinese dominate discussions of phonological tone, but many linguists who look at Korean for the first time, myself included, have no idea anything tonal is going on. Second, even if one argues differently for young Koreans from Seoul, it's for the most part true that there are no minimal pairs based on tone in Korean, unlike in Japanese. The tonal pattern is clearly below the level of consciousness. I asked two Korean linguists, who do not specialize in phonetics or phonology, but nevertheless have much more knowledge on the topic than non-linguists, about the difference between the plain and aspirated stops. Despite a clear tonal difference in their own speech as they repeated the sounds to me, they were not aware of this difference. Thirdly, as noted above, this is not a typical tonal phenomenon, and is perhaps an otherwise unattested manifestation of phonological tone. We may presume that segment-induced tone is an intermediate stage of historically attested tonogenesis, but nevertheless this situation seems otherwise unattested in modern day languages. Finally, the fact that some authors describe the effect as microprosodic, in the purely phonetic sense, seems to have a perpetuating effect in the literature. See Chapter 2 to see how these opposing view points have played out in the literature.

Now, recall that for young Seoul speakers, the plain and aspirated stops may show no aspiration difference at all (see Chapter 3). Given that the words beginning with these stops differ tonally, it suggests that tonogenesis has *already* occurred. However, there are at least two reasons to think that the phonemes in question haven't merged, which makes it difficult to claim tonogenesis has occurred. First, there appears to be a closure duration contrast even in initial position, described in Chapter 3. Second, the phonation contrast between plain and aspirated is maintained AP-internally, at least in spontaneous speech, so that AP-medial but word-initial phonemes maintain a segmental contrast (but no tonal contrast). Nevertheless it seems that the current situation for young Seoul Koreans must be conducive to tonogenesis, and perhaps we'll get to observe the process. More on the potential for tonogenesis in the Conclusion.

1.3 Other Korean Dialects

This dissertation is about the Seoul dialect. My experiments were mainly restricted to Seoul speakers, and my review of the literature mainly ignores non-Seoul dialects. However, it's worth discussing them briefly here. Jun (1996b, 1998) looks in detail at both Seoul Korean and Chonnam Korean, which at an abstract level are very similar to each other. Jun posits the AP for both dialects, and both dialects display the tonal allophony I have described, with L-APs and H-APs realized based on initial segments. The set of L-segments and H-segments is also the same for both dialects. However, the default contours for the AP in Chonnam Korean are LHL and HHL. In addition, tonal assignment is moraic rather than syllabic in Chonnam Korean, which has maintained the vowel length distinction that Seoul Korean has lost.

In addition, at least one Korean dialect has true lexical tone, the dialect of North Kyungsang (Kim, 1997). In short, each word has exactly one H tone, but the location of the H varies substantially.

1.4 Summary

This dissertation is broad in scope, integrating several research threads from the literature as well as different research methodologies to present a unified picture of the unusual properties of Korean described above. First, most of the existing phonetic examinations and phonological analyses are brought together in one place in Chapter 2. The present chapter has presented the major descriptive points, but Chapter 2 provides the details. Second, segment-induced tone is covered in a comprehensive fashion: how it relates to the phonetics-phonology interface and how it may relate to tonogenesis. All previous work on this topic has been incomplete in one way or another, and while my account my also be incomplete, I believe it is the most comprehensive so far. Third, original data is used to present a picture in apparent time of the aspiration contrast in Seoul Korean based on VOT, closure duration, and perception task responses. Silva (2006) is the first apparent time study of this change; this second study has the same goals with a more detailed approach. Fourth, the issues touched upon elsewhere in the dissertation are corroborated with an analysis of spontaneous conversational speech, which to my knowledge has not been done in these research areas. Finally, mainly in the conclusory chapter, I reflect on how the unusual properties of Korean relate to the phonetics-phonology interface and sound change. These speculations are a guide to extending components of this dissertation into future research projects.

Chapter 2

Literature Review

Many of the works described in this chapter cover both phonetics and phonology to a fair degree (which is unsurprising), but I have tried to conceive of this literature review as separating those two threads. Phonetic research is primarily cummulative, and this dissertation is primarily phonetic, fitting into that accumlation of Korean phonetic information. Phonological work, like anything highly theoretial, is often as contradictory as it is cummulative, and this area of Korean phonology is particularly contentious due to the odd features of Korean obstruents. Specifically, the otherwise unattested features of Korean obstruents make satisfactory analysis elusive. Furthermore, this contention within the phonological research seems to encourage more reference to the phonetic research than otherwise might be the case, or so I perceive. This perception in turn encourages me to separate the two threads. While a single work my address both phonetics and phonology, I feel it's more informative to present the phonetic body of work first, followed by the phonological body of work. The former then serves as a point of reference for the latter, as well as a point of reference for the remainder of this dissertation. I'll reserve some phonological speculations for the conclusory chapter.

This two thread approach will be applied to work on segmental properties, while prosody will be discussed in a separate section.

2.1 Segmental Phonetics

The phonetic work is presented in apparent time order, approximately. In other words, the results are sorted by speaker age, not publication date.

2.1.1 VOT, Closure Duration, and Closure Voicing

It seems sensible to group together discussion of the acoustic measures of VOT, closure duration, and closure voicing. Note that this in effect includes aspiration. VOT and aspiration may or may not be different things depending on one's particular definitions, and I will make note of each author's definitions as appropriate. As far as my definitions, I consider VOT to be a specific acoustic correlate, while aspiration is a matter of phonological categorization. For example, one may choose to describe the category of voiced aspirated stops, which might refer to breathy voiced stops. Such stops would have no VOT. But this is simply my choice of definitions.

Some authors define VOT based on the appearance of F2, what I would call F2 Onset Time. Silva (1992) defines VOT as such, and furthermore reports on a third measure called Vowel Lag, which he defines as the greater of VOT and F2 OT. F2 OT in effect excludes any initial breathy voicing, and equates F2 onset with the onset of modal phonation[1]. More on the measure of VOT in Chapter 3; for now, the distinctions are relevant while summarizing the literature.

[1]This is an assumption I don't share. F2 strictly speaking has no delayed onset, merely a delayed appearance in the spectogram, which is difficult to define in objective, consistent terms

Lisker and Abramson (1964)

Lisker and Abramson (1964) in some sense kicked off the research in this area, both for Korean and for cross-linguistic studies. For Korean, they measured the VOT of word-initial stops from three different contexts: isolated words, sentence-initial position, and sentence-medial position. The data comes from one speaker, age and gender unknown. The authors use true VOT, as far as I can tell.

Table 2.1: VOT data from Lisker and Abramson (1964)

	isolated words								
	p*	p	p^h	t*	t	t^h	k*	k	k^h
mean	7	18	91	11	25	94	19	47	126
range	0-15	10-35	65-115	0-25	15-40	75-105	0-35	30-65	85-200
N	15	30	21	16	24	12	16	34	12
	sentence-initial								
mean	7	22	89	11	30	100	20	48	125
range	0-15	15-40	55-115	0-20	15-40	75-130	0-30	35-75	80-175
N	14	28	24	15	21	12	14	35	10
	sentence-medial								
mean	5	13	75	12	22	78	21	44	93
range	0-10	10-20	40-130	0-25	10-45	50-120	10-35	30-65	55-175
N	14	10	23	16	12	10	14	11	10

We can already see one prosodic trend more closely examined in later works: at stronger prosodic positions there is greater VOT. There seems to be a trend among the aspirated stops here of greater VOT in the two utterance initial positions. One can also see the greater VOT for both plain and aspirated velars, as compared to the other two places of articulation. This pattern is observed repeatedly in the literature, and will be discussed again later.

Han and Weitzman (1970)

Han and Weitzman (1970) examined three speakers of Seoul Korean, 2 male and 1 female, of unknown age. The data comes from "1,400 word-tokens" which presumably means isolated words, but this is not stated. The large values of VOT reported would suggest, based on later studies, that the stops were in utterance-initial position. This is the safer assumption when considering change in progress; comparing this data to utterance-internal position would suggest a change based on an inappropriate comparison. I presume the authors use true VOT; although note that measuring F2 OT would also result in values that are larger than other authors' values[2].

Table 2.2: VOT data from Han and Weitzman (1970)

	speaker 1, male								
	p*	p	p^h	t*	t	t^h	k*	k	k^h
mean range	5 0-15	27 15-45	129 80-185	12 5-25	33 15-80	133 85-190	20 13-35	62 40-100	148 95-205
	speaker 2, male								
mean range	5 0-10	20 8-40	105 75-140	8 3-15	23 13-30	107 75-170	27 13-53	42 15-73	136 110-175
	speaker 3, female								
mean range	5 0-10	17 5-33	66 20-95	6 3-15	21 10-40	73 45-130	15 8-28	27 15-50	71 40-110

Cho et al. (2002)

In Cho et al. (2002), four Seoul speakers were examined, all male; we can infer the subjects were born prior to 1950. The nine stops and two fricatives were examined in word-initial position, in isolated words. Despite its recency, the age of the speakers and the utterance type make this study more comparable to the earlier studies described above (assuming age-grading is not at work in the variables under examination).

Only one mean VOT value for the Seoul speakers is mentioned specifically in the text, that for plain stops. The other two values are estimated from the plot given. N = 6 for each stop, therefore N = 18 for each of these means. The three-way VOT contrast is a significant one.

[2] I also can't help but wonder if 1970's technology in effect forced F2 OT measurements, unlike the high resolution and facility of Xwaves and Praat. But this is pure speculation on my part.

Table 2.3: VOT data from Cho et al. (2002)

tense	plain	aspirated
20	72	120

Cho and Keating (2001)

Cho and Keating (2001) examine three Seoul speakers, two males and one female. We can infer the speakers were born in the year range of 1962-1968. This study investigates the three alveolar stops as well as /n/. It's a safe generalization across all three speakers to dichotomize the four consonants into /t/ and /n/ on one hand, and /t^h/ and /t*/ on the other, at least for the correlates relevant to this section. The aspirated and tense stops have longer closure durations, which were measured both articulatorily and acoustically. A clear division between the aspirated and tense stops is not always possible, but tense stops tend to be slightly greater in terms of these measures.

Figure 2.1: Cho and Keating (2001) p. 172 acoustic closure duration

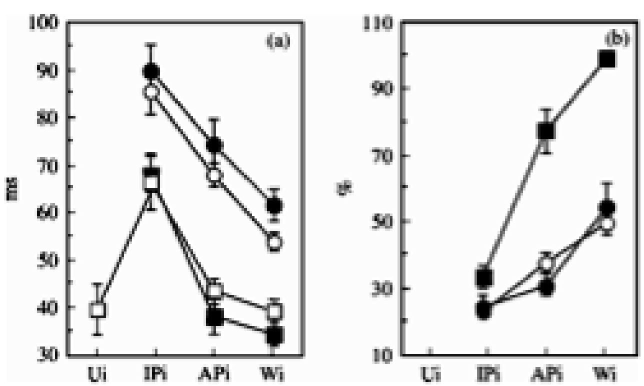


Figure 13. (a) Variation in acoustic duration (ms) for /n, t, t^h, t*/ and (b) variation in %voicing during the acoustic closure duration for /t, t^h, t*/. Error bars show 97% confidence intervals (each datapoint the mean of 60 values). —◆—/t*/; —○— /t^h/; —■—/t/; —□—/n/.

There is a clear difference in VOT, which they define as F2 OT, between plain and aspirated stops at all prosodic positions, which is expected based on the age of the speakers and the results of the other studies. What strikes me as odd about the plot on p. 173 is that the means are in general *lower* than expected, especially given their liberal definition

Figure 2.2: Cho and Keating (2001) p. 173 VOT

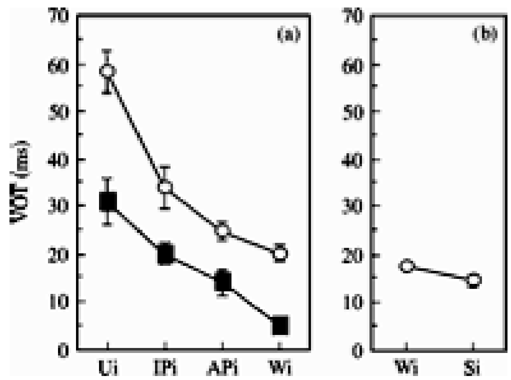


Figure 14. Variations in VOT. (a) higher levels and (b) lower levels. Error bars refer to 97% confidence intervals (each datapoint the mean of 40 values). —○— /tʰ/; —■— /t/.

of VOT as F2 OT. The data points on this plot represent 40 values each, so a paucity of data is not the problem. I can think of no explanation for this; on the other hand, it doesn't necessarily deserve explanation, it is simply curious.

Absolute voicing into the closure does not differ by either manner or prosodic position. Percentage of closure voiced does, because of course the closure duration does. Percentage of closure voiced and total voiceless interval, as derived measures, are not reported here.

Silva (1992)

Silva (1992) examines five male speakers, whose years of birth we can infer are in the range 1954-1973. He examines three different prosodic positions: phrase-edge (PE), word-edge (WE), and word-internal (WI). Here *edge* means *initial*; in addition, both PE and WE positions are utterance-internal. PE is at the subject-object juncture, while WE is internal to the verb phrase.

As mentioned earlier, Silva uses an unusual, derived measure for an effective VOT: he defines vowel lag as the greater of the two primitive measures, VOT and F2 OT. Vowel lag is then the measure which is entered into statistical analysis, rather than either VOT or F2 onset time. Silva notes that vowel lag equals VOT very close to 50% of the time in his data, meaning he saw evidence of F2 before periodic motion about 50% of the time, resulting in VOT being the larger of the two primitive measures. He also reports VOT for backwards compatibility with other studies. The two measures appear to show the same patterns.

This study also reports separate measures for stops that follow vowels and stops that follow nasals. For simplicity's sake, I will not report data for the post-nasal environment.

The following is his ANOVA table for plain stop closure duration (p. 133).

Table 2.4: data from Silva (1992)

place	phrase edge				word edge				word internal			
	cls	vcg	VL	vot	cls	vcg	VL	vot	cls	vcg	VL	vot
p	66	20	67	56	50	20	36	23	48	43	14	5
t	65	15	78	60	53	14	50	30	41	34	12	7
k	56	13	73	70	42	14	46	42	33	27	7	8
p^h	85	15	85	78	77	14	78	71	84	14	47	38
p*	104	20	10	9	81	16	12	10	123	14	12	10

When it comes to the diachronic constrast between plain and aspirated stops (discussed in Chapter 3), the following quote is most relevant:

> Lax and aspirated stops, however, exhibit vowel lags that are longest at PE position, intermediate at WE position, and shoretest at WI position. In each position, the values for vowel lag for the aspirated and lax stops are significantly different, thereby leaving the two stop types distinct from each other. Yet the interaction between phonation and prosody works such that the vowel lag for a PE lax stop is not signficiantly different from that of a WE aspirated stop, and the vowel lag for a WE lax stop is not significantly different from that of a WI aspirated stop (see Figure 5.8). (p. 166)

Therefore the distributions on the whole have a fair bit of overlap, but when position is controlled for, the differences are significant. PE tends to show strengthening effects over WE; for example, tense stops have a longer closure and aspirated stops have longer aspiration at PE position.

Kim (1994)

Kim (1994) examined the VOT and closure duration for 12 speakers, 6 from Seoul, 6 from Pusan, balanced by gender. The given age range for the speakers is 25-35, so subracting that range from 1999, and adding 2 years to the upper bound, gives a year of birth range of 1962-1974. Individual speaker values are not given, but are grouped together by dialect and/or gender. Values are given separately for each place of articulation, but I will report just one place, specifically bilabial, since that is the place of articulation I examine in Chapter 3. Each speaker read each phoneme 6 times (6 different words). Since the following tables show speaker groups of size 3, N = 18 for each cell.

Table 3.11: Mean and Standard Deviation (SD) of Voice Onset Time for Bilabial Stops

		p^h		p		pp	
Gender	Dialect	Mean	SD	Mean	SD	Mean	SD
Male	Seoul	76.7	18.2	45.6	18.2	8.7	3.2
	Pusan	75.1	22.8	14.6	7.1	7.0	1.1
Female	Seoul	71.2	20.7	77.7	18.4	7.8	1.9
	Pusan	87.4	2.5	64.2	15.0	10.9	2.1
Total		77.6		50.5		8.6	

Table 2.5: VOT data from Kim (1994)

		p*		p		p^h	
gender	dialect	mean	SD	mean	SD	mean	SD
male	Seoul	8.7	3.2	45.6	18.2	76.7	18.2
	Pusan	7.0	1.1	14.6	7.1	75.1	22.8
female	Seoul	7.8	1.9	77.7	18.4	71.2	20.7
	Pusan	10.9	2.1	64.2	15.0	87.4	2.5
total		8.6		50.5		77.6	

Table 2.6: Closure duration data from Kim (1994)

		p*		p		p^h	
gender	dialect	mean	SD	mean	SD	mean	SD
male	Seoul	100.9	19.7	63.2	4.1	80.2	6.4
	Pusan	84.1	6.8	71.9	6.4	77.0	10.5
female	Seoul	123.2	30.9	107.9	40.5	128.6	44.3
	Pusan	112.8	11.8	73.6	10.9	91.7	4.5
total		105.3		79.2		94.4	

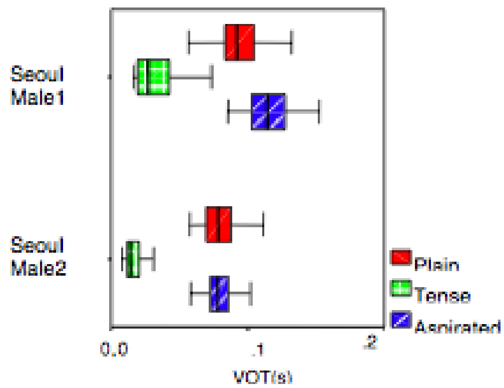
Table 3.4: Mean and Standard Deviation (SD) of Closure Duration for Bilabial Stops

		p^h		p		pp	
Gender	Dialect	Mean	SD	Mean	SD	Mean	SD
Male	Seoul	80.2	6.4	63.2	4.1	100.9	19.7
	Pusan	77.0	10.5	71.9	6.4	84.1	6.8
Female	Seoul	128.6	44.3	107.9	40.5	123.2	30.9
	Pusan	91.7	4.5	73.6	10.9	112.8	11.8
Total		94.4		79.2		105.3	

While reasoning about speakers grouped together in this way can be problematic, it does appear that female speakers from Seoul have lost (or reversed) the aspiration contrast between plain and aspirated stops. At the same time, it seems that all speakers maintain a closure duration contrast between those manners, an important point when considering the possibility of merger.

Choi (2002)

Choi (2002) examined the VOT of all nine stop categories in word-initial position. The target words were nonsense words of the form CVCV, or more specifically, /_apa/, and the context was a carrier phrase that placed the word in sentence-medial position. Three of the speakers were from Seoul, two male, one female. The paper says the speakers were in their late twenties, so if we assume a time span of 25-30 years prior to the date of publication, the speakers were born in the range 1972-1977. Specific measured values and means are not given in the paper. Below are the box plots for the Seoul males. (p. 7)



Voice onset was defined as the onset of F2. The three-way VOT contrast is significant for speaker 1, but there is no contrast for speaker 2 between plain and aspirated stops. No plots are given for the female speakers. For the Seoul female speaker, there is a two-way contrast of VOT, with plain and aspirated not significantly different from each other.

2.1.2 Pitch

The following studies, described above, also measured F0

Han and Weitzman (1970) reports the following F0 values. The values here strike me as high all around, for both genders; this is perhaps due to the authors using narrow band spectrograms, whereas I use the pitch tracker in Praat. Nevertheless the pattern of L-segments vs. H-segments is as expected.

Kagaya (1974), most notable for the articulatory data and discussed below, also measured pitch for two speakers from Seoul. It's unclear how to interpret the /VCV/ data, since the properties of the AP should dominate in that position (at least in real words, which these are not). Furthermore, the values are not entirely consistent with other studies, for example note the low value for aspirated stops for speaker KZ.

Table 2.7: VOT data from Han and Weitzman (1970)

Table II. Onset values of fundamental frequency following a stop consonant (cycles/second)

	Range ave. min. max.	Range ave. min. max.	Range ave. min. max.
Informant 1			
	/pʰ/185 150.0–220.0	/tʰ/205 175.0–233.0	/kʰ/201 166.0–230.0
	/p/ 144 130.0–170.0	/t/ 161 130.0–173.0	/k/ 162 140.0–170.0
	/P/ 178 145.0–215.0	/T/191 177.0–225.0	/K/188 170.0–205.0
Informant 3			
	/pʰ/341 317.5–368.7	/tʰ/334 307.5–367.5	/kʰ/343 310.0–389.0
	/p/ 266 250.0–292.0	/t/ 312 292.5–327.5	/k/ 309 292.5–327.5
	/P/ 308 277.5–322.5	/T/345 325.0–370.0	/K/337 320.0–355.0

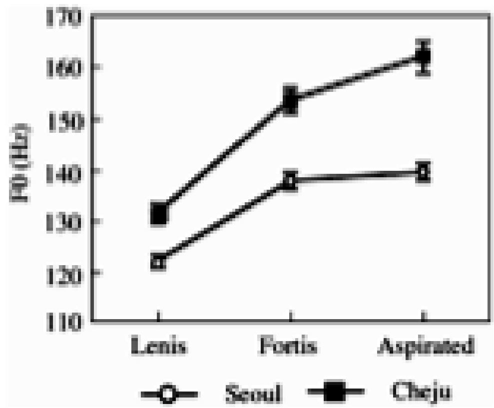
Table II Average value [Hz] of voice onset f_0. The letter "N" represents samples for each type

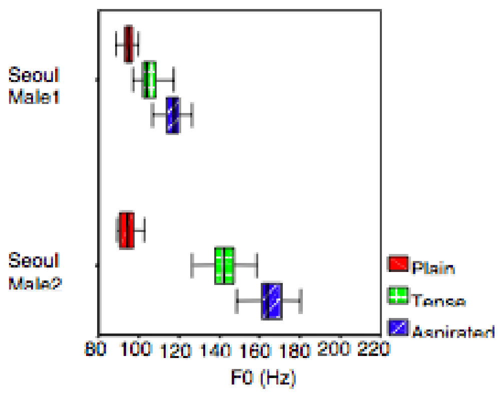
		Forced		Lax		Aspirated		
		Avg.	Range	Avg.	Range	Avg.	Range	N
Subject K.T.								
/CV/	stop	160	(167–148)	148	(160–138)	162	(174–154)	12
	affricate	157	(167–148)	144	(148–138)	157	(160–154)	4
	fricative	165	(167–160)			157	(167–148)	4
/VCV/	stop	159	(167–143)	141	(154–133)	150	(160–138)	12
	affricate	149	(160–138)	138	(138–138)	147	(154–143)	4
	fricative	151	(160–143)			145	(154–133)	4
Subject K.Z.								
/CV/	stop	192	(210–174)	162	(182–148)	150	(160–138)	12
/VCV/	stop	192	(200–182)	166	(182–154)	167	(182–154)	12

Cho et al. (2002) find that F0 is significantly lower for plain stops than the other two stops, when pooled together across all other categories, including dialect and position-in-vowel. The pattern seems equally valid for Seoul speakers alone, however, based on the given plot.

Choi (2002) measures F0, describing the measurement as follows.

> Fundamental frequency value was taken from the onset at the phonation phase after the stop release with the pitch analysis of Praat program, using autocorrelation function. The onset value was taken as the mean of five initial values, which were measured every 0.01 second from the beginning of the second formant. (p. 6)





The three-way F0 contrast is significant for both speakers, with aspirated stops having higher F0s than tense stops. There is a three-way significant contrast for F0 (presumably in the same order as for the male speakers).

Silva (2006) shows F0 values in terms of syllables rather than true time. Focusing on the first syllable, the difference between L-segments and H-segments is clear.

2.1.3 Other acoustic measures

What follows is a sample of other measures from the literature, which I have given less weight to overall. While interesting on a purely phonetic basis, these other measures, I believe, are less important for broader issues like contrast. For example, some authors have measured the intensity of stop bursts and/or aspiration. However, it's unlikely that

Figure 2.3: F0 values from Silva (2006)

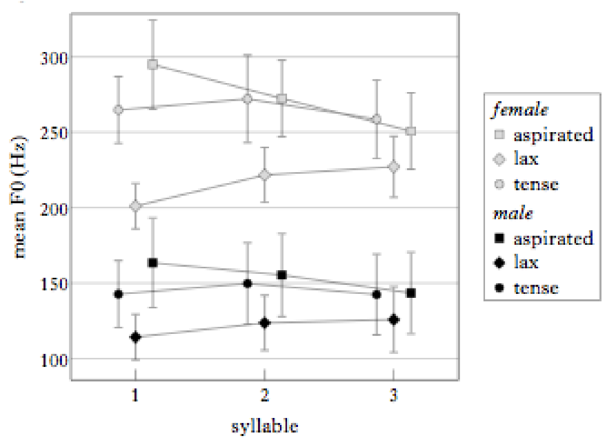


Figure 4
Comparison of averaged tonal melodies of male *vs.* female speakers. The phonation types indicated refer to the first C in each target word. The bars represent 1 SD above and below the mean value.

aspiration intensity, at least, varies independently from VOT, given work that connects VOT to glottal width (Kagaya, 1974; Kim, 1970). Furthermore, it's unheard of in phonology for a language to contrast segments in terms of stop burst or aspiration level.

Harmonic intensities are perhaps more important to investigate, but I simply haven't given full thought to this correlate yet.

Cho et al. (2002)

Cho et al. (2002) also discuss burst energy and harmonic intensity. For burst energy comparisons, only one significance value is given, so it's unclear whether the other patterns discussed are in fact significant ones. When all speakers of both dialects are combined, the aspirated stops have a significantly greater burst energy than the other stops, while plain and tense stops show no difference. The Seoul speakers together seem to follow this pattern; however, note that a higher burst energy of plain stops over tense stops

> is in fact observed for only two out of four Seoul speakers and five out of eight Cheju speakers, suggesting that the observed difference in the figure [regarding

> Cheju speakers] is due to speaker variation regardless of Dialect, ... (p. 14)

The authors are emphasizing the probable lack of a true dialect difference, whereas I am emphasizing the point that burst energy patterns may be speaker dependent; although, aspirated stop burst energies seem to be consistently higher.

In the discussion of H1-H2 results, perhaps the most important statement is the following one.

> All four Seoul speakers make a clear three-way distinction among stops, showing a pattern of fortis $<$ aspirated $<$ lenis in H1-H2. (p. 16)

Or in my terminology, tense $<$ aspirated $<$ plain. The authors describe how this is indicative of tense, aspirated, and plain stops having creaky, modal, and breathy phonation respectively at voice onset. The differences are apparent at the vowel midpoint, but to a lesser degree. For H1-F2 comparisons, tense stops are significantly lower than other stops, even at vowel midpoint, which indicates more abrupt closure. (need to say more here)

2.1.4 Articulatory measures

Cho et al. (2002)

Cho et al. (2002) find that intraoral airflow values follow this hieracrhy: aspirated $>$ plain $>$ tense. They find that intraoral airpressure values follow this hierarchy: (aspirated=tense) $>$ plain.

Cho and Keating (2001)

Cho and Keating (2001), whose acoustic measures where discussed above, also examined articulatory closure duration and percent linguopalatal contact via an electropalatograph.

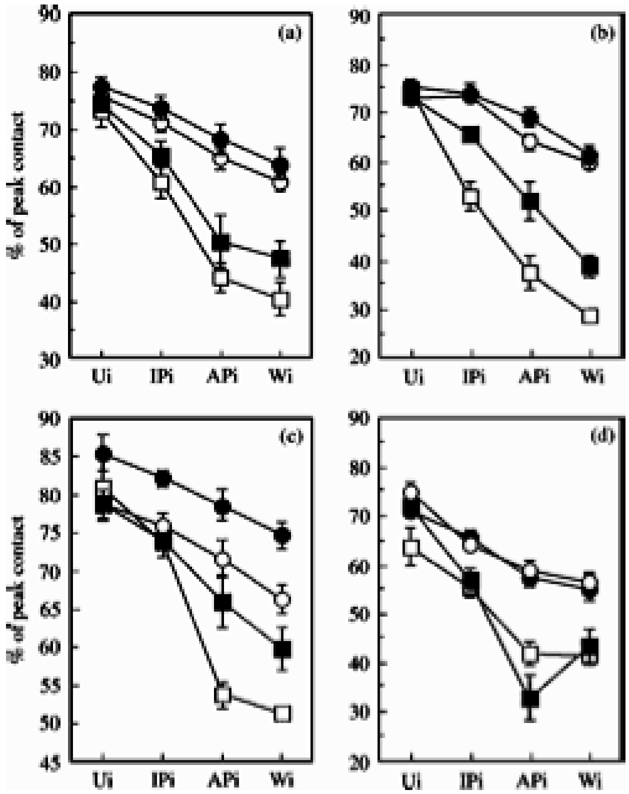


Figure 6. Peak contact for each consonant over the entire pseudo-palate area. (a) contact across speakers (each datapoint the mean of 60 values); (b) contact for speaker NHL (each datapoint the mean of 20 values); (c) contact for speaker THC (each datapoint the mean of 20 values); and (d) contact for speaker JYY (each datapoint the mean of 20 values). Error bars show 97% confidence intervals. —●—/t*/; —○—/tʰ/; —■—/t/; —□—/n/.

There results, acoustic and articulatory, show a robust pattern of cummulative prosodic strengthening. Of interest here, however, is primarily manner contrast. There seems to be at least a dichotomy between L-segments and H-segments in terms of both these articulatory measures.

Kagaya (1974)

Kagaya (1974) uses a fiberscope to film the glottis during the articulation of the Korean obstruents. The results show a clear three way contrast of the three manners. In initial position, for tense stops, the glottis begins slightly open, but is adducted from well before stop release to immediately after release. For intervocalic tense stops, the glottis behaves

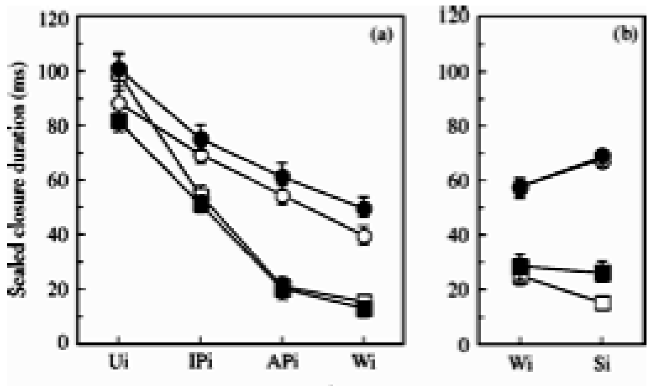


Figure 9. Seal duration as a function of prosodic position and consonant type. Data pooled across speakers (each datapoint the mean of 60 values). Error bars show 97% confidence intervals. (a) higher levels; (b) Wi *vs.* Si. —●— /t*/; —○— /t^{h}/; —■— /t/; —□— /n/.

similarly: around the time of closure, the glottis opens slightly before returning to fully adducted position for the portion of closure preceding release. For initial position plain stops, the glottis is abducted during closure; intervocalically, the glottis is adducted for voicing. For aspirated stops the glottis is widely abducted in both positions, having a similar width to that of the glottis in respiratory position. In initial position, the aspirated stops have a substantially wider glottis than the plain stops. What's particularly interesting about Kagaya's measurements is that both plain and aspirated stops, while having different glottal widths during closure, both have a constantly adducting glottis from the point of stop release. Acoustically, since aspirated stops have a longer VOT (for many speakers), it suggests that the adduction gesture is delayed when compared to plain stops. However, Kagaya's data shows that adduction occurs at stop release for both plain and aspirated stops, and that longer VOT can be attributed to wider glottal width.

the glottal width actually tends to be adducting before release in the case of plain stops.

Hirose et al. (1974)

Hirose et al. (1974) conducted an electromyographic study of the laryngeal muscles during Korean stop production. Of particular interest is what they found for tense stops, when

Figure 2.4: Initial stops, Kagaya (1974)

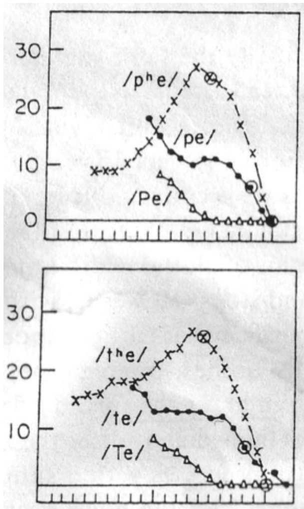
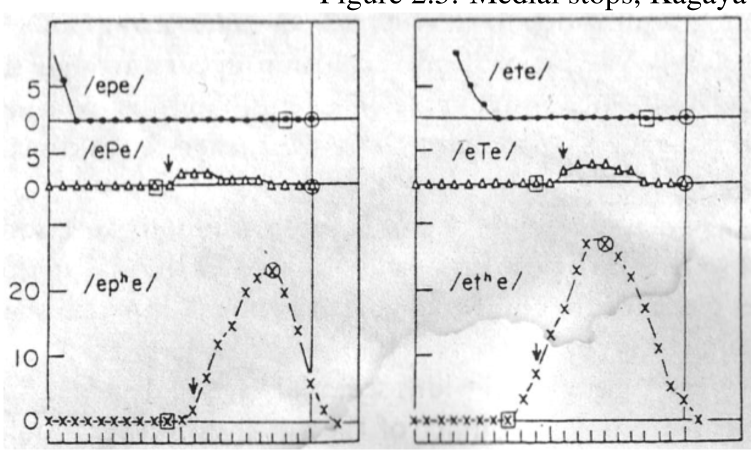


Figure 2.5: Medial stops, Kagaya (1974)



compared to the other two manners. They found that the vocalis and lateral cricoarytenoid (LCA) muscles demonstrate high activity for the tense stops, but not for plain and aspirated stops. The vocalis muscle is the muscle forming the core of the vocal folds themselves, and its activity, its *tensing*, lends empirical support to other claims of the tense nature of tense segments. The LCA is one of the two major adductor muscles of the glottis. The other adductor muscle, the interarytenoid muscle, did not show activity for the tense stops.

2.2 Segmental Phonology

There is a wide variety of phonological descriptions in the literature on Korean obstruents. Martin (1951) is responsible for an early multi-segment analysis: this analysis treats aspirated segments as a sequence of plain stop plus /h/, and tense stops as plain stops plus /q/, a glottal segment of some sort. However, the vogue since (at least) Kim (1965) is to use single segment, feature matrix analyses. Numerous features, both privative and binary, have been proposed to capture the three way manner contrast, including [tense], [stiff], [slack], [spread glottis], and [constricted glottis] (along with traditional features like [voice] and [aspiration]). Kim (1965) and Kim-Renaud (1974) both use a binary feature [tense], which is phonetically motivated by the apparent tenseness of the vocal folds and/or vocal tract walls, but is nevertheless somewhat abstract, just as [aspiration] is.

		stop manner		
	features	plain	aspirated	tense
Kim	[tense]	-	+	+
	[aspirated]	+	+	-
Kim-Renaud	[tense]	-	+	+
	[aspirated]	-	+	-
	[voice]	-	-	-

Kim's account doesn't require an additional feature to capture voiced plain stops, which would be [-tense,-aspirated]. However, this is arguably a drawback, since it's natural to think of an intervocalic voicing rule in terms of a [voice] feature. Kim's account captures the fact that plain stops do have aspiration in initial position, but only by requiring [+tense] to contribute to aspiration in the case of aspirated stops, when it clearly doesn't in the case of tense stops.

Kim-Renaud's account is somewhat better in these respects, except that it doesn't capture the fact that plain stops actually have some aspiration. The advantage of having aspirated stops [+tense] is to capture the fact that aspirated and tense stops share properties like higher pitch.

Halle and Stevens (1971) offers a more detailed feature system, using [spread], [constricted], [stiff], and [slack] as laryngeal features. Note that while [spread]/[constricted] and [stiff]/[slack] are pairs of opposites, the features are all binary.

	plain	aspirated	tense
[spread]	+	+	-
[constricted]	-	-	+
[stiff]	-	+	+
[slack]	-	-	-

This account is similar to Kim (1965) in that the aspiration of plain stops is captured, while relying on [+stiff] to add to the aspiration of aspirated stops. Although, at least one can argue that [+stiff] doesn't add aspiration to the tense stops by virtue of them being [+constricted]. The deconstruction of features like [tense] and [aspirated] into more specific articulatory features is an advantage, although the explosion of combinations is potentially a disadvantage. Since [+spread,+constricted] and [+stiff,+slack] aren't possible, there are nine possible types, which the authors treat one by one. This is an interesting account of laryngeal operation, however it is specifically flawed in its treatment of Korean. Begin with this quote:

> Kim has observed that the fundamental frequency in the vowel following the release of a stop tends to be higher for the voiceless stops [p] and [p^h] than for the partially aspirated [p_k].

The notation is not a typo on my part: for Halle and Stevens (1971), [p] is my /p*/

(tense), and [p_k] is my /p/ (plain). For Halle and Stevens (1971), like Lisker and Abramson (1964), the tense stops of Korean are simply voiceless unaspirated stops, explicitly that which is seen cross-linguistically. In turn, the plain stops are the typological rarity: Korean is used as the illustration for this stop manner. This point deserves more discussion, but for now just note the results of Kagaya (1974): tense stops have adducted vocal folds, and are not the average voiceless stop.

Many accounts analyze tense stops as geminates. One attraction of this type of account is that nothing about the tenseness of tense stops needs to be specified underlyingly. It's further supported by the phonetic evidence, since word-internal tense stops have a longer closure than word-initial ones. Silva (1992) pursues such an approach, representing tense stops as plain stops with two X-slots, which is the condition that results in phonetic tenseness. Initial tense stops de-geminate due to syllable structure contraints. (what about ordering?)

Silva (2006) in part retreats to the intuition of Kim (1965) and Halle and Stevens (1971) that tense and aspirated stops form a natural class. He posits a privative [stiff], which is underlying in aspirated stops, and derived in tense stops, which are again represented as geminates. This way, the two manners are a natural class at the point of phonetic implementation, specifically for the purposes of tone assignment. Aspirated stops are also said to have [spread] as an underlying feature, while the younger speakers of Seoul in addition add [spread] to word-initial plain stops, accounting for the equal or close to equal aspiration observed for those speakers. Silva's goal is to provide a uniform underlying representation for the Seoul community despite the phonetic variation by age that he observed.

See Ahn and Iverson (2004) for another geminate account of tense stops. The authors cite other studies that support the existence of initial geminates, which in turn support their analysis of Korean. However, it's likely that the other languages cited have initial geminates which are actually longer versions of the singletons. In Korean, tense stops are not simply longer versions of plain stops.

2.3 Prosody

2.3.1 The Accentual Phrase

The introductory chapter summarized Jun's theory of the Accentual Phrase (Jun, 1996b, 1998). Her theory is an alternative to previous syntactic accounts of phonological phrase definition (Silva, 1992; Kang, 1992). In her account, an utterance can be chunked into APs in different ways "depending on speech rate and other non-syntactic factors" (Jun, 1996b, p.66). The APs are then assigned dialect specific tonal contours, subject to the H-AP/L-AP variation already discussed, and serve as the domain of application for several phonological rules. Although AP placement is highly variable, Jun does give these rules (Jun, 1996b, p.187):

every prosodic word may be an AP a focused word must be the left-most word in an AP an ap can include any numnber of prossodic words as long as: i. the last prosodic word is not theleft welement of a branching constituetn ii. all the prosodic words are not ocused.

It seems that the only morpho-syntactic domain which must be a single AP is a stem plus its suffixes. All other boundaries can be aligned to AP boundaries.

2.3.2 Microprosody and Tonogenesis

The microprosodic effects of stop manners have been well established, if also somewhat contested in their details (Cristo and Hirst, 1986; Silverman, 1986). The broadest generalization is that voiceless stops perturb F0 upwards and higher than voiced stops. This supports a theory of tonogenesis where microprosody is phonologized: voiceless stops create high tones and voiced stops create low tones (Hombert et al., 1979; Kingston and Diehl, 1994).

The connection of this area of research to Korean becomes confusing when reading the literature. Jun (1996b, p.46) stresses that segment-induced tone is not microprosody, since it's non-local, and furthermore, it doesn't obey the standard microprosodic pattern: all Korean stops are voiceless. At the same time her claim is that microprosody has been phonologized to produce the tonal pattern. In actuality, she is not being contradictory, she simply does not use the word microprosody for *automatic* phonetic effects, even though some authors do use it that way (Cristo and Hirst, 1986). A discussion of Jun's analysis is warranted, but not within this chapter. For now, suffice it to say that in her analysis the tone is phonological, and it is connected to the featural analysis of Halle and Stevens (1971): in that analysis the H-segments are [+stiff].

Consider the following two quotes from Cho et al. (2002, p.220) which presumably reflect Jun's co-authorship of the paper. The two quotes are combined here with an ellipsis into one.

> Jun (1996a) claims that the substantial difference in f0 between the lenis and other stops cannot be understood simply in terms of the phonetic pitch perturbation caused by the voicing of the preceding consonant (i.e., microprosody) as found in many other languages (Hombert, 1978; Hombert et al., 1979). ... This phonologization of microprosody is confirmed by our data from Seoul, ...

The contradiction may be due to the multiple authors. It's clear that the first use of microprosody refers to *automatic* perturbations. The second use reflects Jun's actual usage of the word.

Consider the following quote from Cho et al. (2002, p.221).

> ... it seems safe at the moment to conclude that the acoustic perceptual information of the preceding consonants spreads to the following vowels to a large

extent ...

This immediately follows a discussion of f0 differences. In my mind, this quote reflects a reluctance on the authors' part to accept the presence of phonological tone, instead of vowel quality spreading from the consonant and manifested as harmonic intensity differences. Their measurements of harmonics may be completely accurate, but that shouldn't obscure the fact that H-APs have a high tone associated with the first syllable (at the surface level), and the larynx is manipulated accordingly.

I agree whole-heartedly with the quote from Cho et al. (2002, p.222).

> It would take a great deal of procrustean effort to force Korean stops into the categories that have been developed for phonological descriptions of other languages.

And I've never even heard the word *procrustean* before. But I get the gist. The analysis the authors eventually put forward is one of phonological privatives and redundancy rules. Tense segments have the feature [constricted] and aspirated segments have the feature [spread]; the plain segments are laryngeally underspecified. Redundancy rules then add the feature [stiff] to segments that are laryngeally specified. This appropriately accounts for the similarities and differences between the segment types. It also has the advantage of explaining plain stop voicing by relying on passive voicing, due to their underspecified nature. Recall that plain stop voicing seems to be variable and sensitive to prosodic factors.

An opposite approach is to *not* introduce new features, and rely on the more typical underlying contrast of voiced and voiceless unaspirated stops, plus voiceless unaspirated stops. For example, Kim and Duanmu (2004)[3] offer a detailed analysis where the tense stops are underlyingly plain, and the plain stops are underlyingly voiced, subject to an

[3] This type of approach is not new; see references therein. In particular see Kingston and Diehl (1994); Korean is discussed in the footnotes on p. 435.

initial de-voicing rule (aspirated stops are just that underlyingly). The analysis is based on the tonogenesis theory: H-segments are voiceless and L-segments are voiced (in this particular analysis). Whatever feature is used for voicing, for example [+/-stiff], that feature conditions the vowel tone. Plain segments then devoice. Segmental phonetic differences can then be explained by using the tones themselves (and initial position) as conditioning context.

However, tone aside, this sort of approach commits one to an odd system. Underlying plain voiceless stops are never realized as such, always surfacing as tense, besides codas where all stops are realized as voiceless and unreleased. The plain voiceless stops which are realized in initial position, are instead due to underlying voiced stops. Furthermore, since tense stops have no special underlying features, their properties are purely a matter of phonetic implementation. This is passing the buck to phonetic universals. Tense stops are unique, period. Removing the need for a phonological [tense] feature underlyingly just means Korean has a unique phonetic mapping procedure.

Chapter 3

Spontaneous Speech

To complement my elicited laboratory data, I examined some spontaneous Korean speech. The Korean Telephone Conversations corpus (Ko et al., 2003a,b) published by the Linguistic Data Consortium contains 44 hours of transcribed conversational Korean. It's important to note that the audio in this corpus was recorded as part of the CALLFRIEND collection at the LDC around 1996, therefore when giving a speaker's year of birth, I have subtracted their age as given in the documentation from that year. To identify speakers I have used their 4 digit IDs from the corpus, with the addition of the traditional m or f at the beginning to indicate gender.

This chapter includes numerous waveforms, spectrograms, and pitch tracks of corpus utterances. I've defined *utterance* in terms of the transcripts from the corpus: each entry in the transcripts is an utterance, each containing a pair of timestamps that delimit the utterance. A subset of the transcript entries, containing all examples discussed here, is included as Appendix A. I refer to utterances by the integer second at which they begin, rounding down, since these numbers serve as an index into the transcript. All spectrograms are presented here at a fixed width, so are not to scale relative to each other. Unless otherwise noted, the spectrogram for a particular utterance begins and ends at exactly the times given in the transcript. The Praat cursor(s) is used to mark or delimit portions of spectrograms, and should be visible as faint vertical lines.

The discussion in this chapter mostly concerns pitch and tone, with only some mention of segmental properties. The first section discusses aspiration. The goal of the second section is to corroborate the basic facts about tone, in other words, to confirm that the expected patterns appear in spontaneous speech. The third section looks at possibly unexplored issues, particularly tonal alignment.

3.1 Aspiration

It's much more difficult to reason statistically in spontaneous data than laboratory data since the former is not controlled. A typical one half-hour conversation doesn't provide that much data, considering the variation it presumably represents. Furthermore, there just aren't many word-initial aspirated stops, compared to plain stops. To illustrate the difficultly of this, I provide a histogram below of word-initial obstruents, plus /h/. This histogram is an *average* histogram: I calculated the histogram for all the corpus transcripts together, then divided the values by 200. There are 100 conversations in the corpus, therefore 200 people.

Table 3.1: Average histogram for word-initial obstruents, plus /h/

	manner		
series	plain	aspirated	tense
p	42.5	5.3	2.5
t	58.4	5.5	17.2
k	215.6	3.3	4.5
c	81.2	13.4	3.3
s		53.0	5.0
h		68.2	

The following table shows some VOT measures for speaker m6129. The two speakers mentioned in this section, m6129 and f6753, are focused on in later sections. They were born in 1973 and 1954, respectively.

Table 3.2: VOT values for speaker m6129 (1973)

p	42 67 75
t	37 48 59 61 97 128
k	25 28 41 52 55 58 59 63 64 65 72 87 87
p^h	38 68 79
t^h	51 59 80
k^h	50 56 92 117 135

The predominance of /k/ is in part due to the determiner /ki/. The large value of 128 for /t/ appears to be emphatic, when heard in context. And, as said elsewhere, we expect velars to have longer VOTs than anterior places.

There certainly appears to be overlap of plain and aspirated distributions, but the data is too sparse for statistical analysis. There are more measurements from this data set, but they are primarily more cases of plain segments, so comparison by manner is still too difficult. Also consider the following VOTs, all from the initial segment of Toronto, which is /t^h/: 29, 46, 46, 48, 51, 55, 56, 61; mean = 49.

To give one more example, here are some measurements from speaker f6753.

Table 3.3: VOT values for speaker f6753 (1954)

p	76 77 85
t	67
k	17 28 30 35 36 39 44 46 48 53 110
p^h	29 50 51 92 96
t^h	64
k^h	49 56 100

This speaker, based on her age, would certainly show a clear contrast in careful speech, but the contrast isn't really evident from these numbers.

This corpus is theoretically a great source for examining the change in question, or other changes, but it doesn't seem that meaningful data on aspiration can be extracted.

The amount of data for each speaker is too small considering the great variation that these variables show in natural speech.

3.2 Tone and Speaker m6129

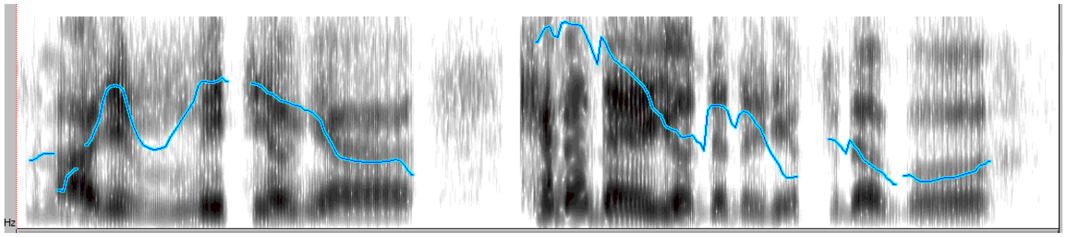
I examined speaker m6129, born in 1973, in detail with the goal of corroborating some basic facts about Korean phonetics and phonology. Of primary interest was the nature of AP-internal H-segments and whether or not they affected pitch. In short, they do not, and the claim that H-segments only trigger H tone AP-initially is supported in this data.

I chose this particular speaker in part for some interesting cases of borrowings into Korean. The following place names occur in this conversation: Toronto, Canada, Quebec, Montreal, and Philadelphia, some more than once. The names appear to be pronounced as Korean, not code-switched English. English aspirated stops tend to be borrowed into Korean as aspirated stops, and interestingly the /t^h/ and /k^h/ from Toronto, Canada, and Quebec all trigger high tone, as would be expected from those segments. This is scant evidence on its own, but it suggests a future experiment, particularly for young speakers: how do they phonologize words actively borrowed into Korean? If speakers begin to borrow English aspirated stops as Korean plain stops, that would suggest a true merger (or initial neutralization) of plain and aspirated stops.

3.2.1 APs and segment-induced tone

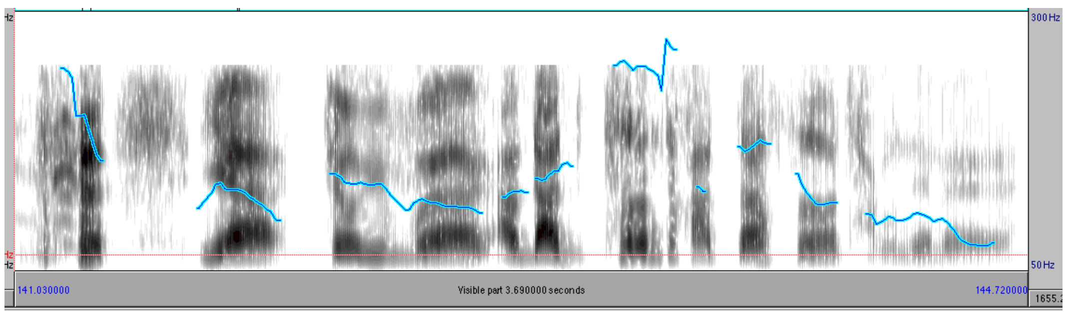
The first few examples are meant to illustrate APs, particularly H-APs, to give some idea of the impact of segment-induced tone on the pitch contour. The multi-word nature of APs is also of interest, since that's for the most part not evident in laboratory speech.

Utterance 122



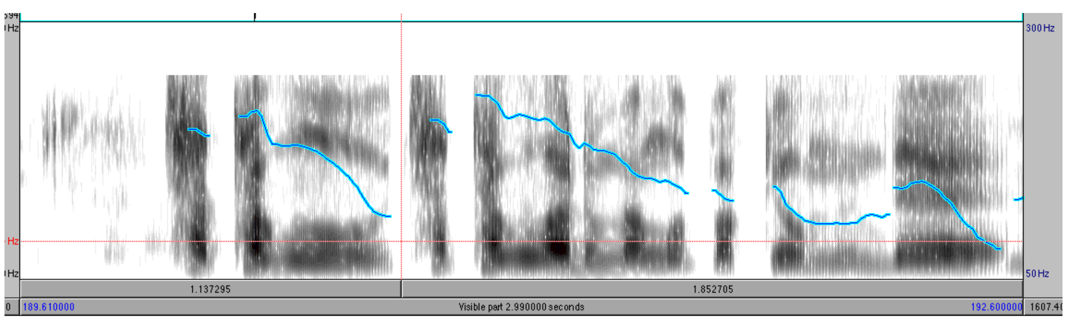
Canada is the first word after the pause in the center, and is responsible for the large pitch peak.

Utterance 141



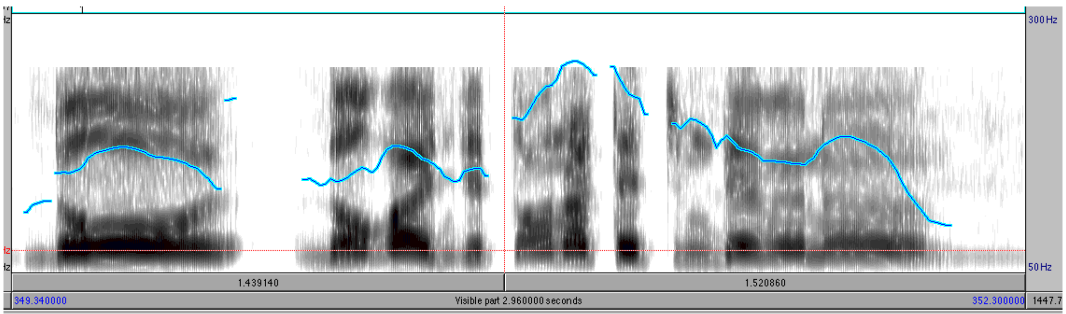
The elevated pitch at the beginning of this utterance is due to a laugh, but the elevated pitch near the center is due to the word Toronto.

Utterance 189



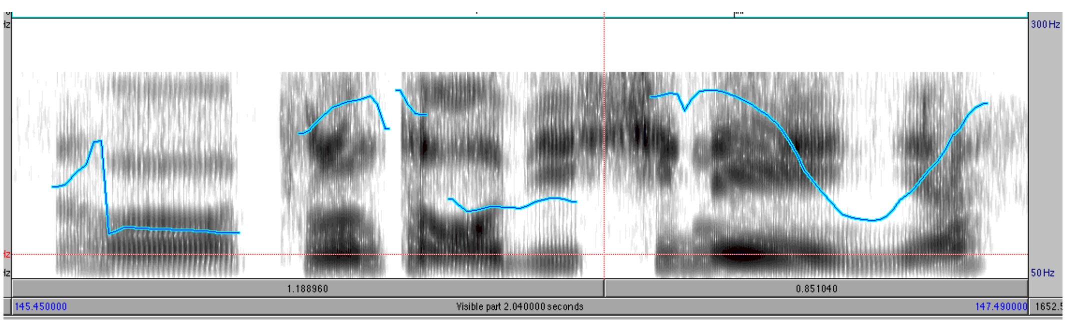
This utterance contains a restart, with the two words /cha tha-ko/ occurring on both sides of the cursor (in other words, the restart involved a repition of the two words). The portion after the restart, after the cursor, is a single H-AP, with a HHLH contour. The two tense stops in the word /pang-pap-pakk-e/, one derived and one underlying, do not affect the downward trend of the contour. This utterance also suggests that H-APs have a second position rise.

Utterance 349



The second half of this utterance forms a single AP, with three tense segments: tta-nal ttae ccim twi-myan. The second and third tense segments don't affect the contour. There again seems to be a second position rise.

Utterance 145



The first syllable here is a filled pause. The rest of the utterance is comprised of two clear H-APs. The first begins with Toronto, and the second begins with the segment /s^h/, marked by the cursor and clearly visible in the spectrogram.

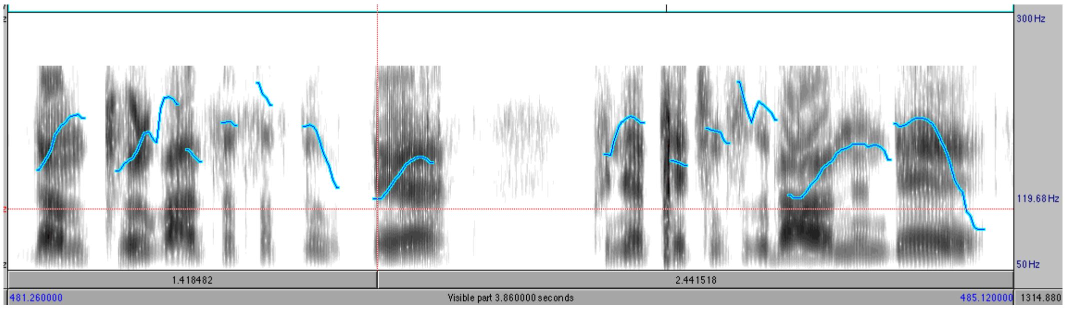
3.3 AP-internal Segments

Tense and aspirated segments can occur word-internally, as well as word-initially while still being AP-internal. Both manners can also be derived from plain stops, by hypothesis only in AP-internal position. One reason for looking at spontaneous data is to confirm that none of these instances of H-segments trigger high tone, that segment-induced tone is only an AP-initial phenomenon. Note that by definition of the AP, derived H-segments cannot occur in initial position.

Furthermore, the segmental properties of derived segments are of interest, to confirm that they match their underlying counterparts.

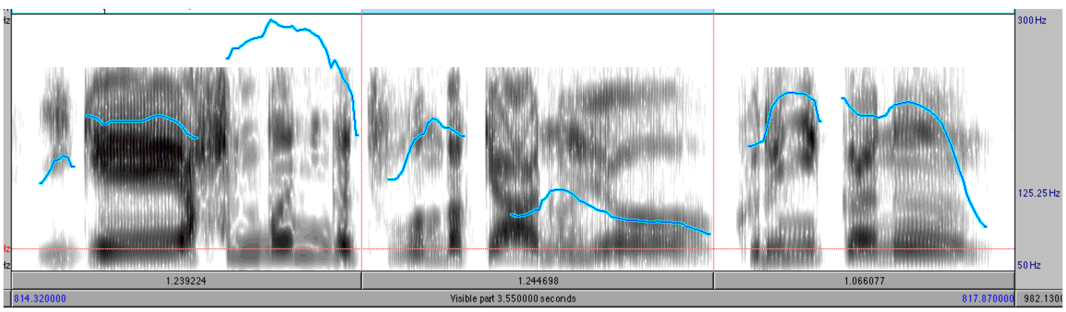
3.3.1 Underlying Tense Segments

Utterance 481



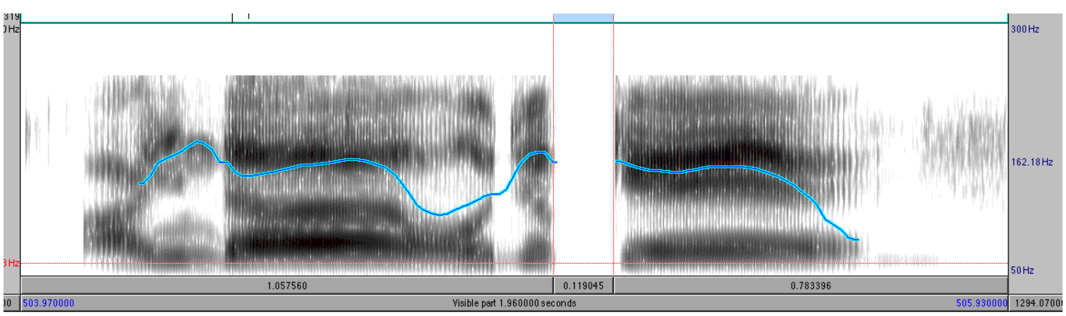
Here the cursor marks an underlying /t*/ which falls right at the penultimate L of the AP.

Utterance 814



Here an L-AP is delimited; the gap in the center is an underlying /k*/ which also falls at a penultimate L.

Utterance 503



The delimited gap here is an underlying /k*/, which immediately follows the second position high of the L-AP.

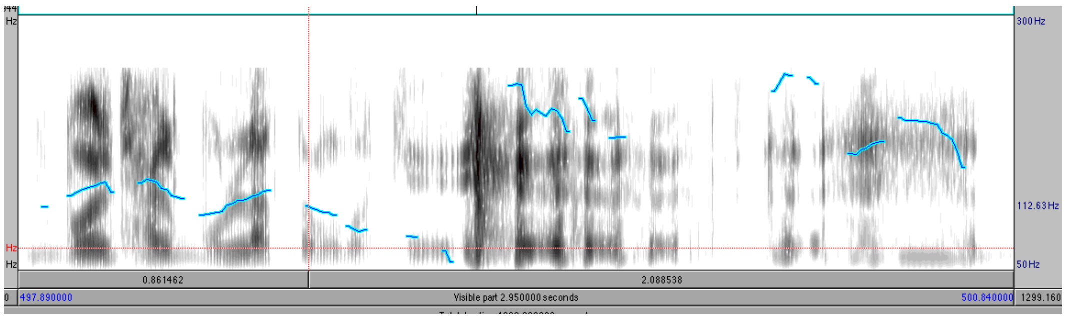
3.3.2 Underlying Aspirated Segments

No good examples that aren't in first or second position of the AP.

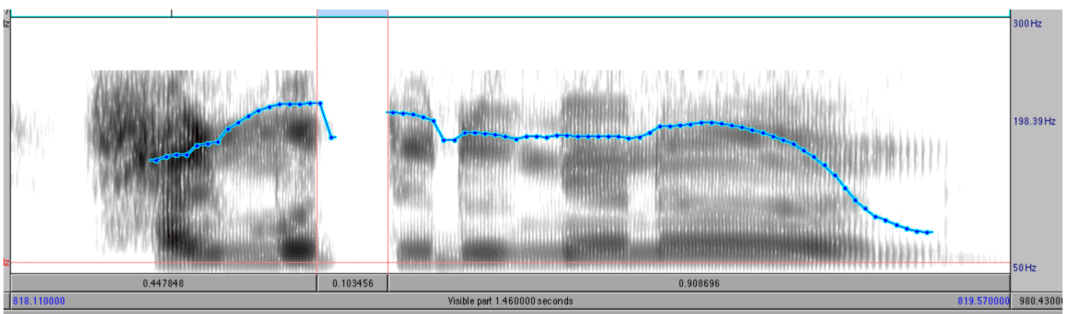
3.3.3 Derived Tense Segments

The following three examples mark derived tense stops, /t*/, /t*/, and /k*/, respectively.

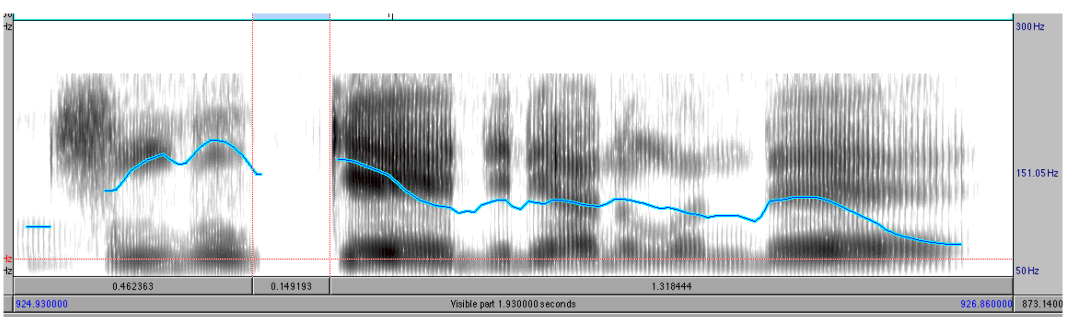
Utterance 497



Utterance 818



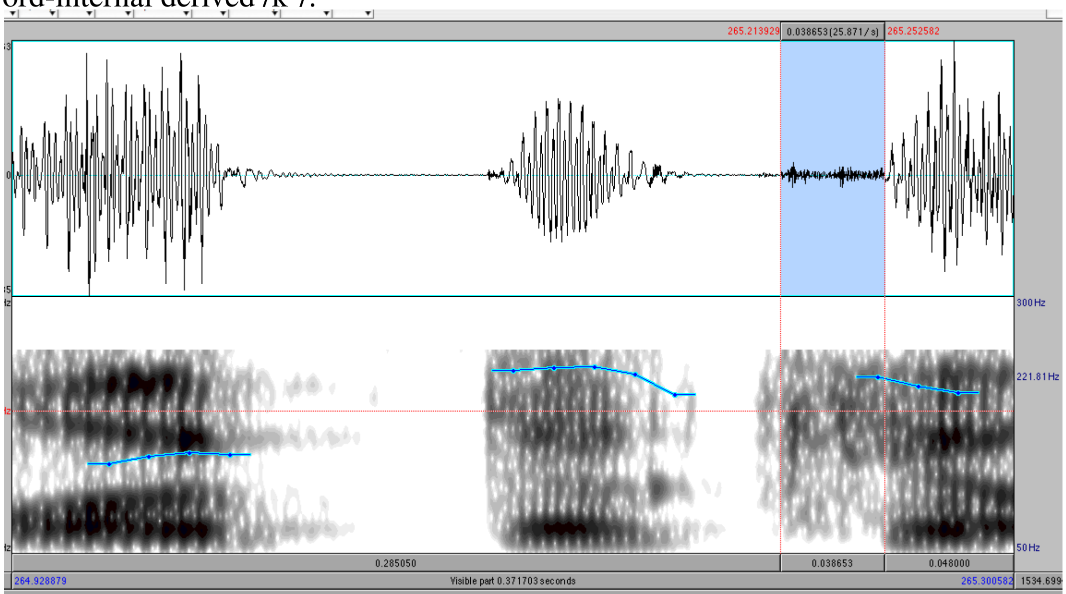
Utterance 924



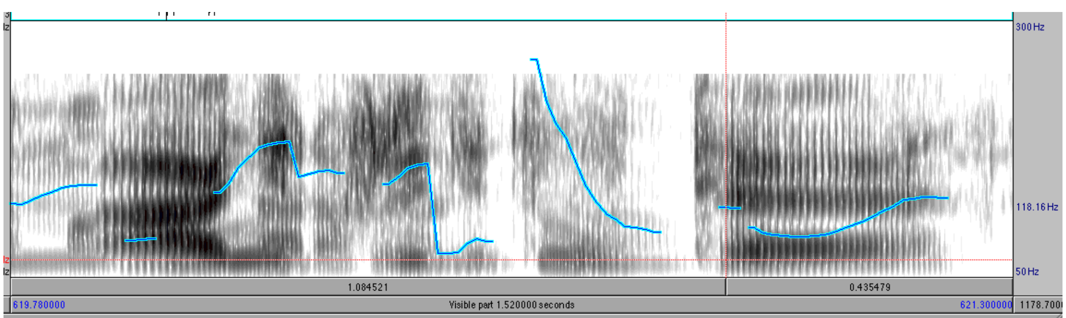
3.3.4 Derived Aspirated Segments

Utterance 261

This picture is a close up of the word o-ttoh-ke (the first of two instances) which illustrates a word-internal derived /k^h/.



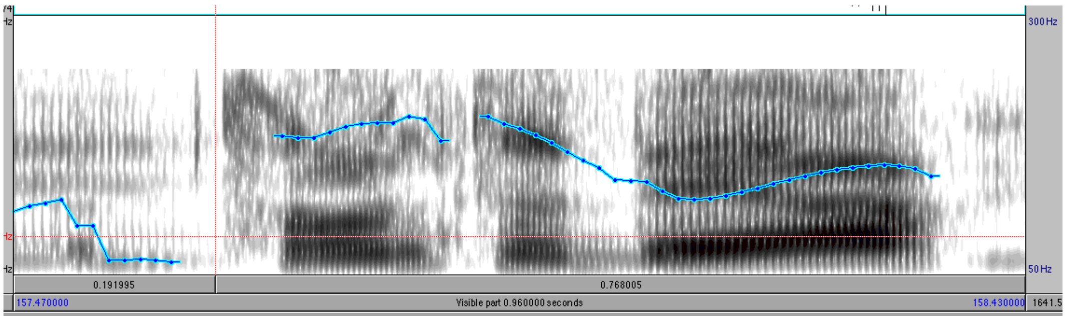
Utterance 619



Here the cursor marks a derived /t^h/ at the word boundary of the sequence *mos hae*. Remember that coda /s/ becomes /t/, feeding the aspiration rule.

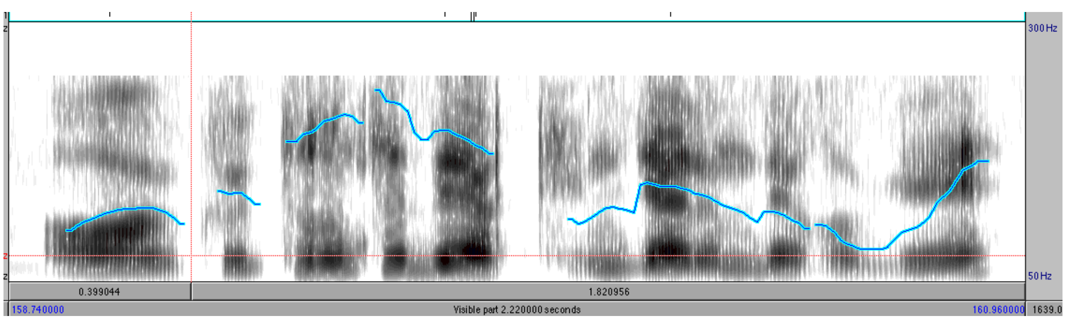
3.3.5 Plain Segment Voicing

Utterance 157



This utterance, beginning at the red vertical line, says Toronto tae-hak, or Toronto University. The /t/ in tae-hak is clearly voiced, showing that plain stop voicing can occur in word-initial position even for content words. It does appear that this phrase is a single AP, especially based on the gradual fall of pitch in the center of the phrase, rather than an abrupt transition from H to L.

Utterance 158

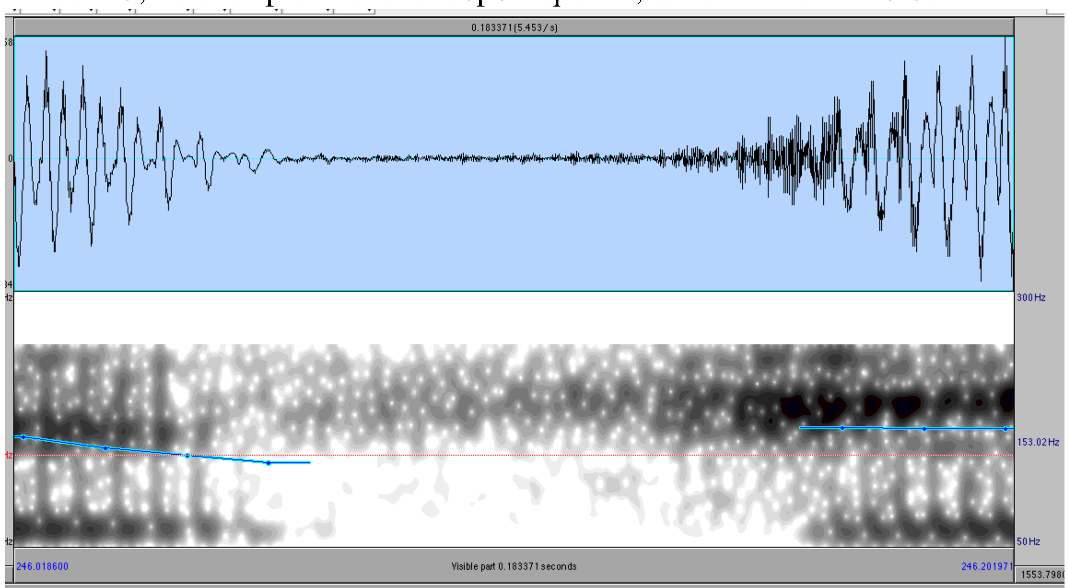


Here it appears that Toronto tae-hak forms a single AP, with the following words forming a second. The /t/ in tae-hak is again voiced.

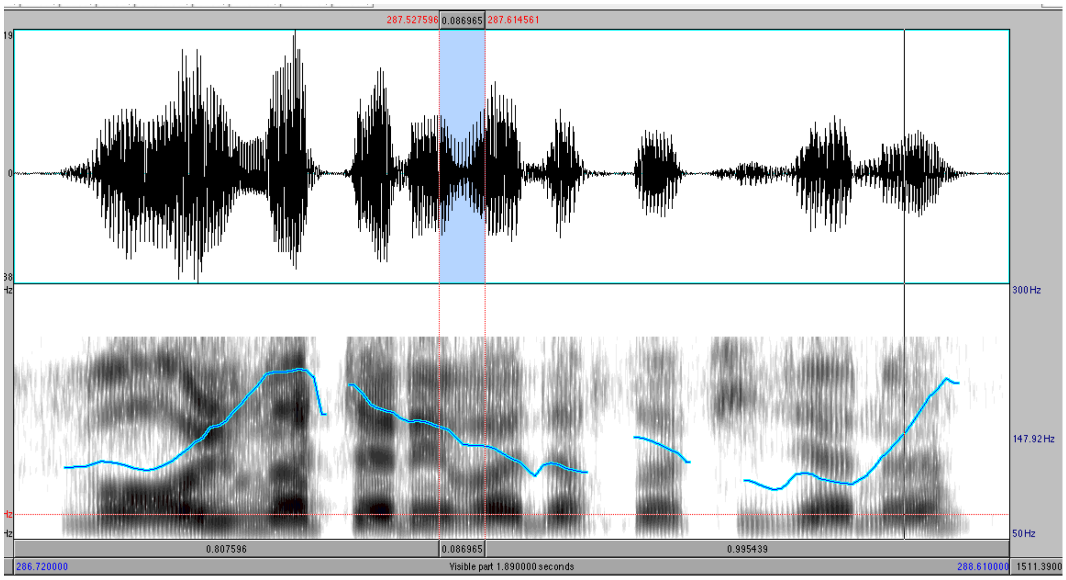
Also of note here is that the determiner *ki* does not appear to be included in the initial AP, based on the high starting pitch of Toronto. The status of small prenominal modifiers like this is interesting in that they do not tend to join APs in laboratory speech, but I presumed they would in fast speech.

3.3.6 Plain Segments

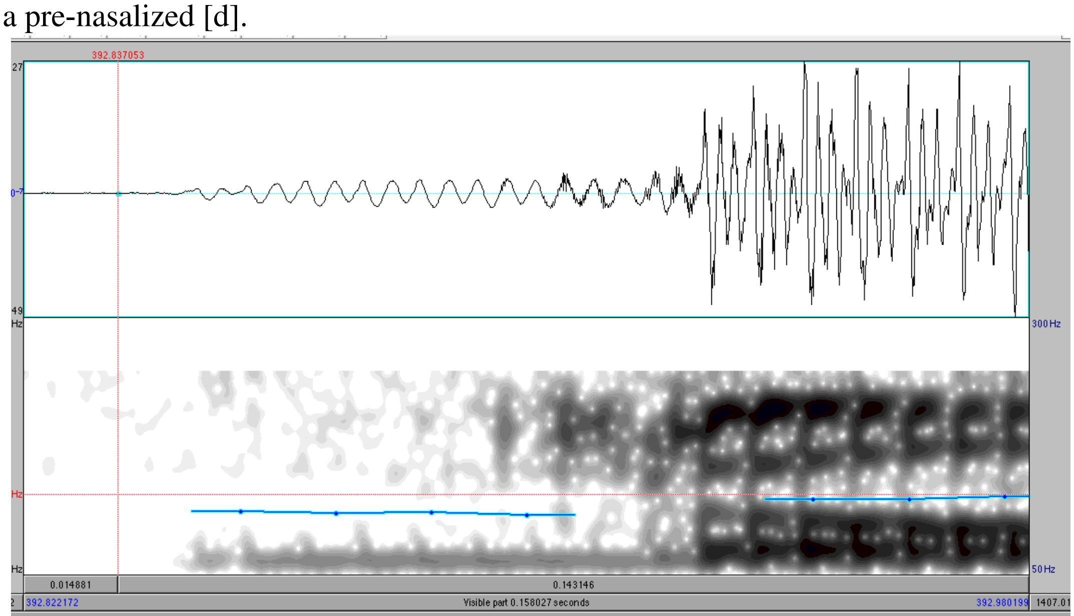
Plain segments are the most malleable segments, becoming tense, aspirated, or voiced in AP-internal position. Furthermore, there are a fair number of plain stops that sound like fricatives or approximants, which has been reported before, at least for velars (Silva, 1992). This is perhaps to be expected since Korean has no underlying labial or velar fricatives, so fast speech spirantization has no functional impact, presumably. Here's and example from Utterance 245, a close up of the second /pi/ sequence, which sounds like /fi/.



Utterance 286, shown in its entirety, has a /p/ that sounds approximated (selected in spectrogram).



In my laboratory data, I often got the impression that the segment /n/ was pronounced as [d], or perhaps a pre-nasalized [d], in initial position. This is a very interesting phenomenon since of course there are no underlying voiced segments initially. Sure enough this seems to occur occasionally for speaker 6129, as well as at least one case of /m/ sounding like [b]. The following close up, taken from Utterance 392, is a good example of what appears to be a pre-nasalized [d].



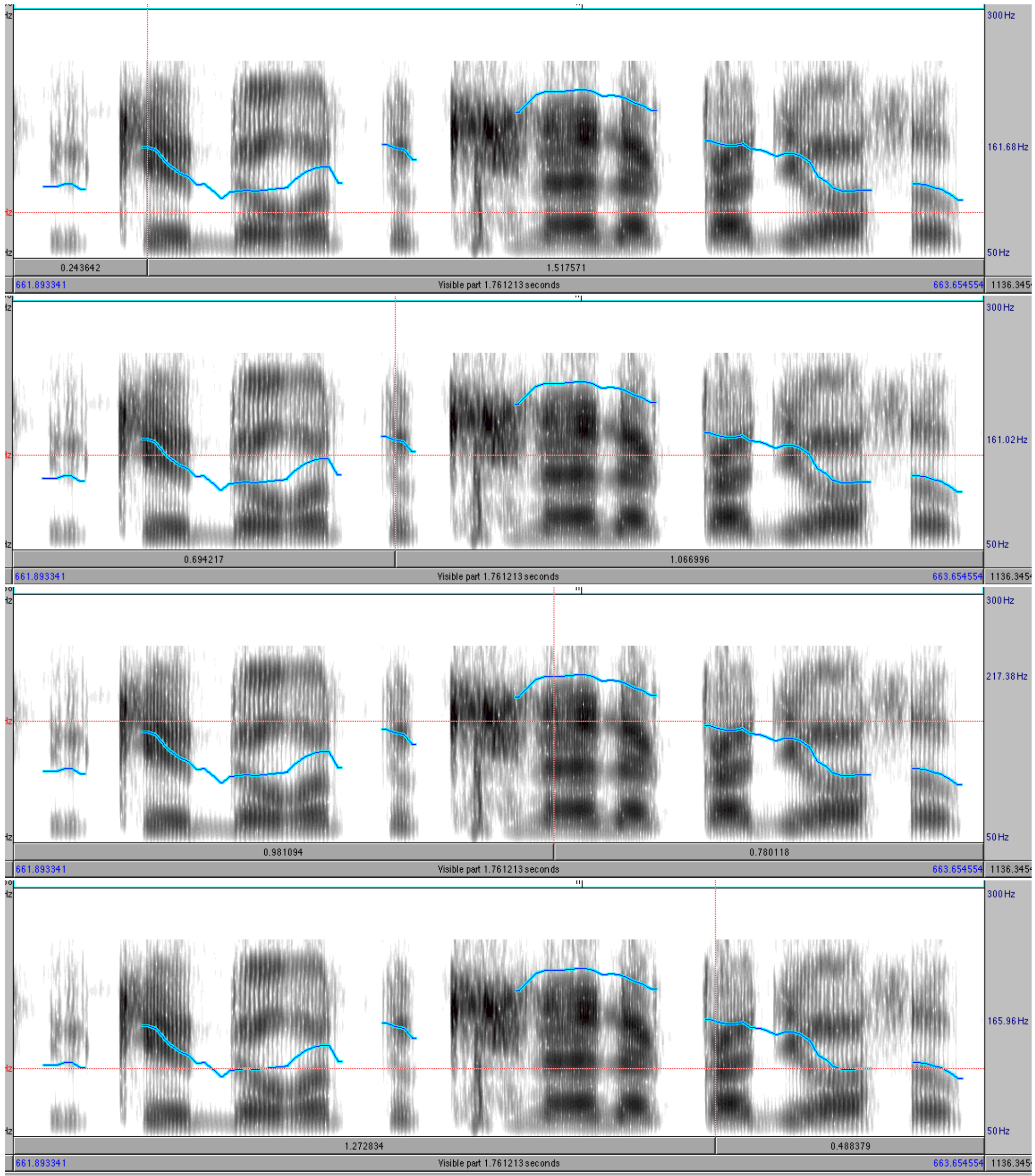
3.3.7 Particles

Consider this selected portion from Utterance 659: i-lah-ke na-ka tto Canada-kka-ci wa-sa. The following 4 spectrograms show the same AP, but with different points in the pitch track: the derived /k^h/, the underlying /t*/, the second syllable of Canada, and the underlying /k*/.

The derived /k^h/ appears to lie at the second position H of the AP, despite being in the third syllable; the measured pitch value is 162 Hz. It's difficult to tell whether the word *tto* is part of the preceding AP or forms its own. It depends how lax we can be with tone alignment, because the pitch is rising well before *tto*. The measured pitch value at that point is 161 Hz. The first vowel of Canada is mostly devoiced, so the pitch is measured at what must be the second position H. However, the high level of this tone must be due to the initial segment of Canada being /k^h/. The measured pitch value is 217 Hz. The final pitch measuremeant of 166 Hz follows the underlying /k*/, but is part of the falling contour of the AP beginning with Canada.

In general, for this speaker, second position H tones of L-APs are around 160 Hz, while the first and second H tones of H-APs are well over 200 Hz, even as high as 250 Hz.

See the next section for more on the particle *tto*.



3.4 Potential New Issues

3.4.1 Speaker 6753

I began looking at the older speakers in the corpus for potential contrasts to younger speakers, and chose to examine speakers f6753 and f6762 in detail, born in 1954 and 1956 respectively since their speech was relatively clear.

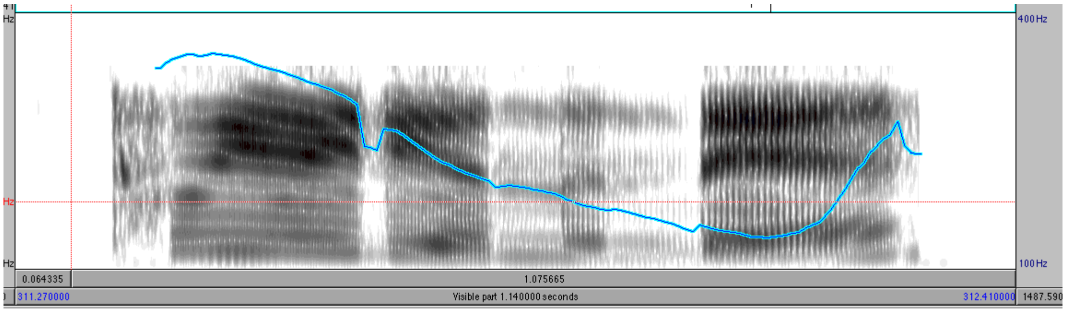
One thing I was concerned with was the fact that some speakers appear to have AP contours of LHL rather than LHLH. Note that according to Jun (1996b), LHL is the default pattern for the Chonnam dialect. Therefore it's not unreasonable that this pattern might show up in Seoul speakers. Perhaps some segment of the community, like older speakers, have a LHL pattern. Or perhaps this is simply a fast speech phenomenon, where the final H of the LHLH pattern is not realized.

APs that end an utterance or precede a pause often have boundary tone effects. For example, in my data collection, the verbs, which were sentence final, usually were lowered overall, and had a contour more like LHL, with no final H. Alternatively, there is sometimes a sudden drop at the end of an AP such that the contour is more like LHLHL. With speaker f6753, the pattern seemed to be LHL most of the time with a final H as a boundary tone.

This issue is subsumed by the general issue of tonal alignment. Part of my reasoning for the above statements is the following assumption: a pitch minimum in the final vowel indicates an L associated with the final syllable. In laboratory speech, tones do seem to align well with syllables, and this is Jun's theory for Seoul speech. For Chonnam speech, see uses the mora as the TBU. In the corpus data, where I primarily looked at Seoul speakers, it often did not seem true that tones aligned with syllables. I recall at least one example that seemed to show aligned to morae: the first syllable's coda was a nasal, after which the pitch began to drop. This would be a second position high under a mora analysis.

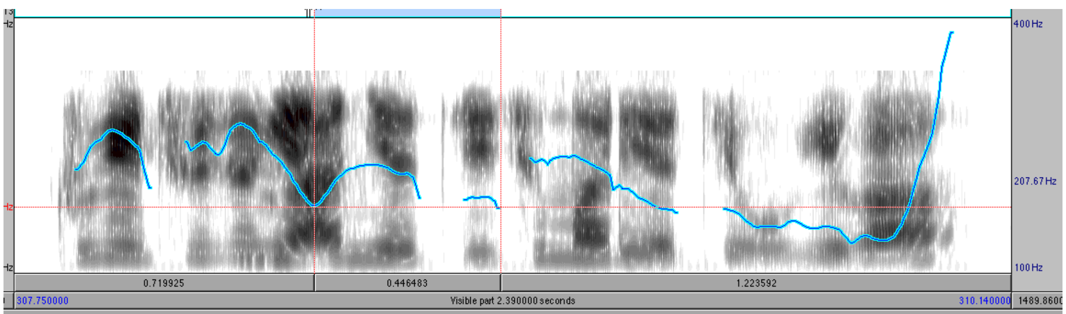
Regardless, my examination here is not comprehensive. The following examples simply suggest that the alignment issue should be considered more in the future, using sponataneous speech.

Utterance 311



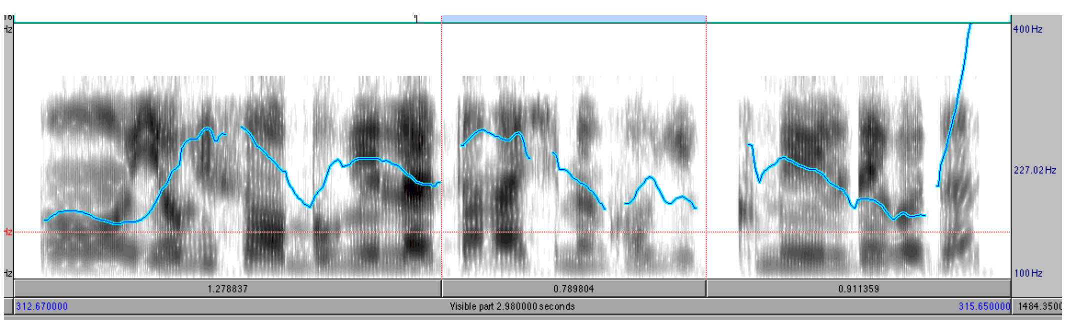
Consider Utterance 311, a single H-AP, which basically just has a HL contour. The final rise that can be seen is so late that it seems to be due to a boundary tone. The sentence is a declarative, or so it seems from the transcript.

Utterance 307



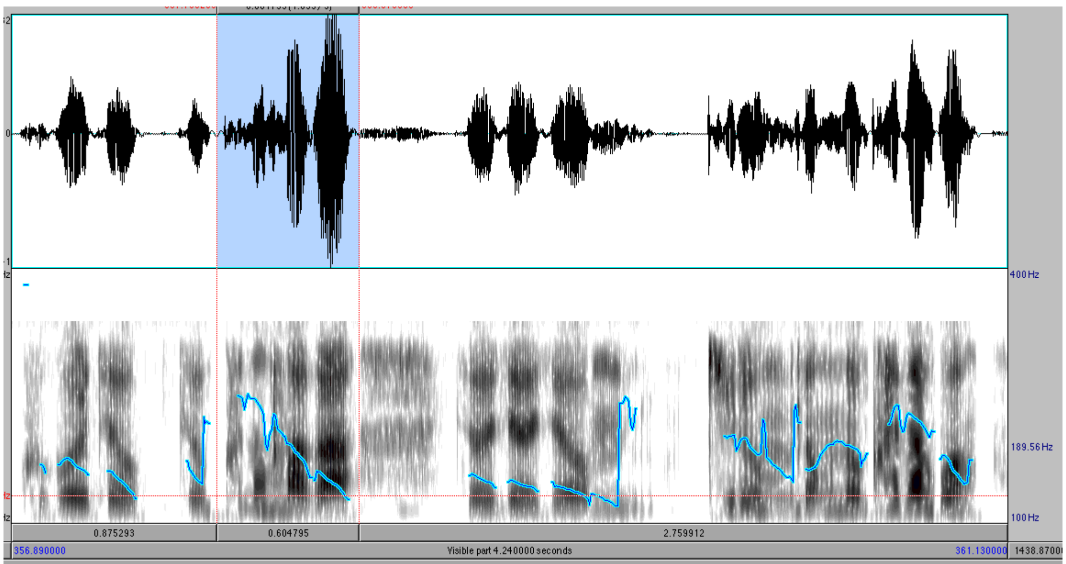
Utterance 307 shows a similar pattern. The selected region is an L-AP, that ends low, and following that there is an H-AP that again seems to end low before an H boundary tone.

Utterance 312



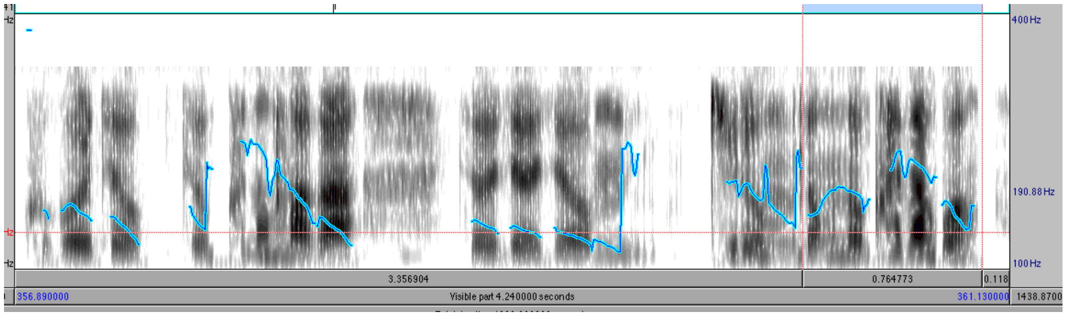
In Utterance 312, the selected region covers two words, tta-na-si-myan ko-ki. This region could be a single AP with the contour HHLH, but the final H appears too early. Plus, the initial /k/ of ko-ki is not voiced. It seems that the words form two separate APs that both end low. Certainly the first AP ends low; the second one does appear to have two pitch peaks, but it's likely that the dip in the middle of ko-ki is due to the microprosody of the stop.

Utterance 356

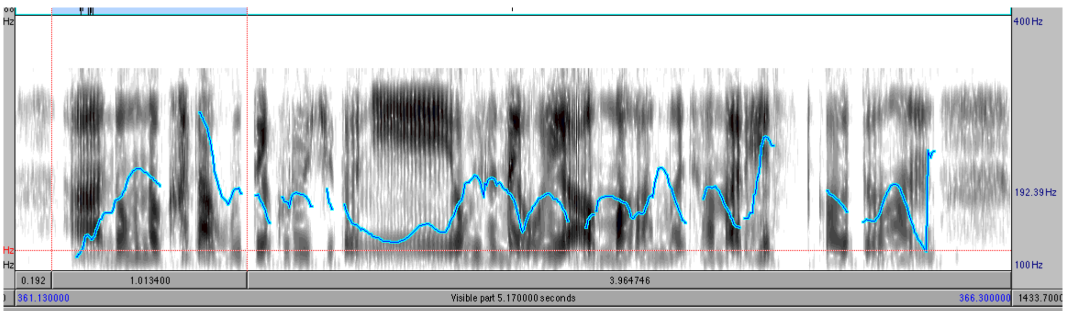


This selection from Utterance 356 clearly shows an H-AP with a steadily falling contour, although one could perhaps argue that the following pause has introduced a boundary tone. The selection covers the word khin-ap-ma-ka.

This same utterance contains a portion interesting for different reasons. The following selection is orthographically at least a single word, nak-nak-ha-ta-ko, although the /h/ clearly triggers an H tone. There is a morphological boundary at that point, so this is probably a case of a single word containing more than one AP.



Utterance 361



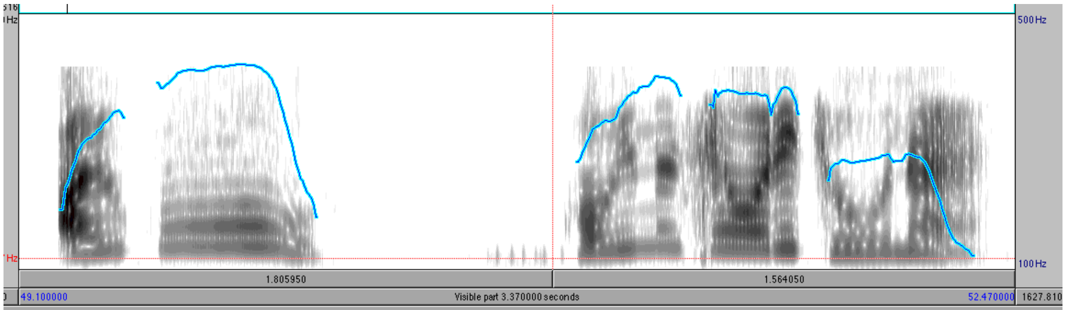
The same word occurs in Utterance 361, shown here.

For this speaker, plain obstruents in AP-initial position seem to be higher than sonorants, although not as high as true H-segments. It's hard to tell if this is systematic or not, but it's possible that older speakers have a different pattern than the the simple division into L- and H-segments which seems correct for younger speakers.

3.4.2 Speaker f6762

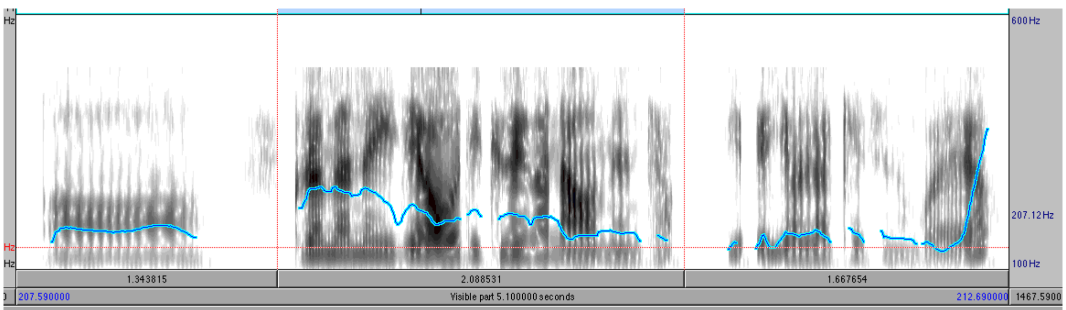
Speaker f6762 has some interesting examples. Her speech was particularly even and metrical.

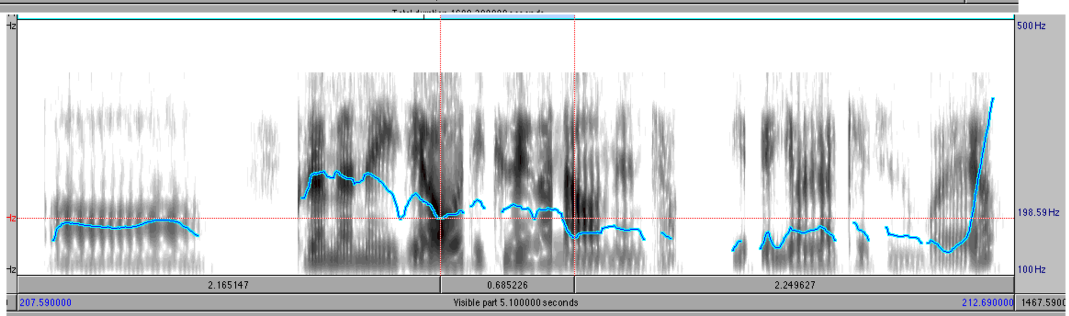
Utterance 49



In Utterance 49, the 4 words a-i-ko ka-ki-nin se-wal-i coh-i-ne are neatly divided into 4 APs. These APs have particularly level contours, which is to be expected when there are only 3 syllables.

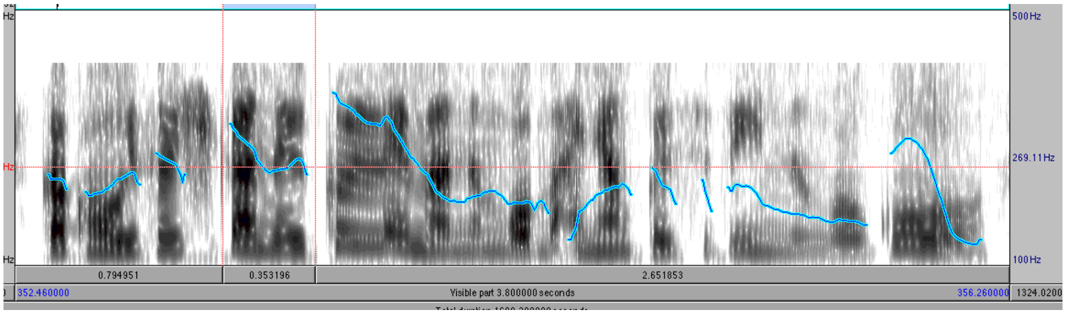
Utterance 207





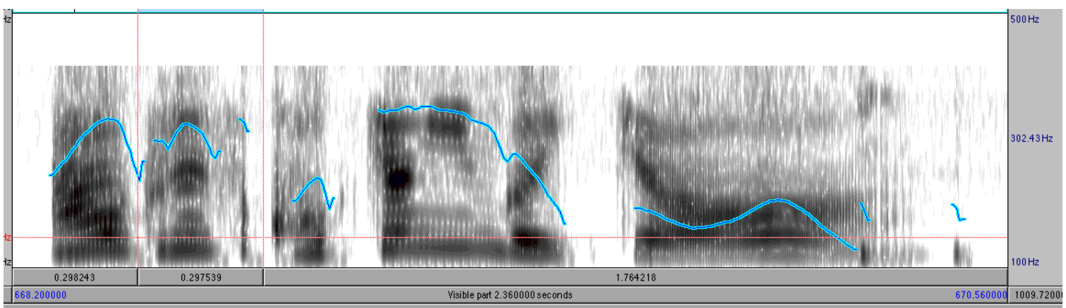
Utterance 207 has what seems to be a clear case of downstepping. The first picture delimits a sequence of three L-APs, while the second picture delimits just the middle of the three, so that the end points of all three are clear. The L tones, as well as the more or less level contours of each AP, noticeably decrease from the first to the second, and from the second to the third. It's quite clear when you listen to the utterance as well.

Utterance 352



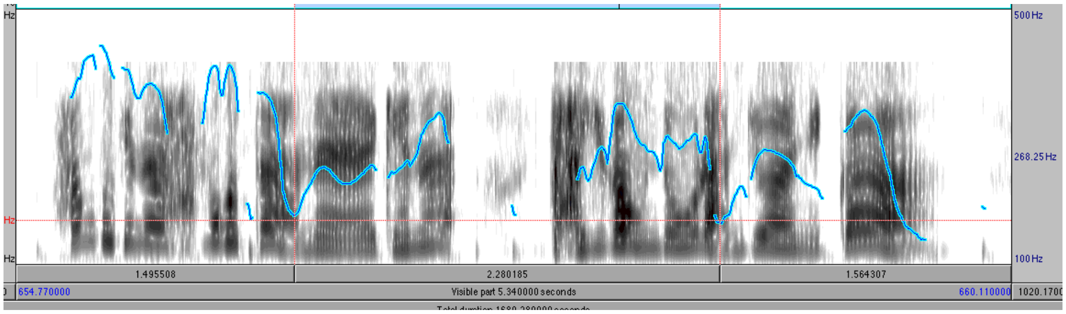
Utterance 352 seems to have an example of upstep. The picture delimits an H-AP, beginning with /s^h/, which is immediately followed by an H-AP beginning with /k*/, the latter noticeably higher than the former.

Utterance 668



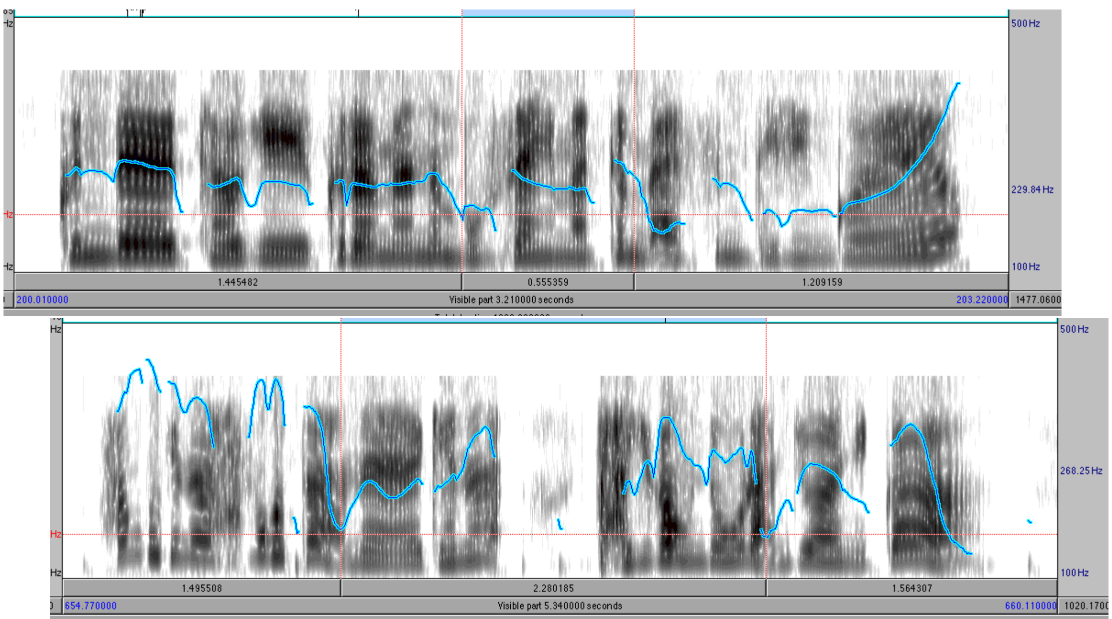
It often seemed the case with this speaker that only sonorants triggered low tones, and that plain stops were basically mid. Descriptively, this is certainly the case for many examples. The delimited AP in this utterance is the word *ki-lae-sa*, but doesn't seem to begin low. However, the immediately following word, *kak-cang-in*, is clearly an L-AP, which suggests that plain stops are still L-segments for this speaker. The question is why many plain stops don't appear to trigger L tone. One possible explanation is that function words don't form true APs. Many function words in Korean begin with /k/, like *ki-lae-sa*, which would add to impression that plain stops don't trigger an L. There may be strong positional or emphatic effects as well, where only prominent words realize a the L tone.

Utterance 654



In this utterance the cursors mark two prominent local minimas in the pitch contour. Each one is associated with the word *an*, the negation morpheme. This word may receive prominence normally, or perhaps the speaker has emphasized those words here; either way it seems the prominence lowers the L tone.

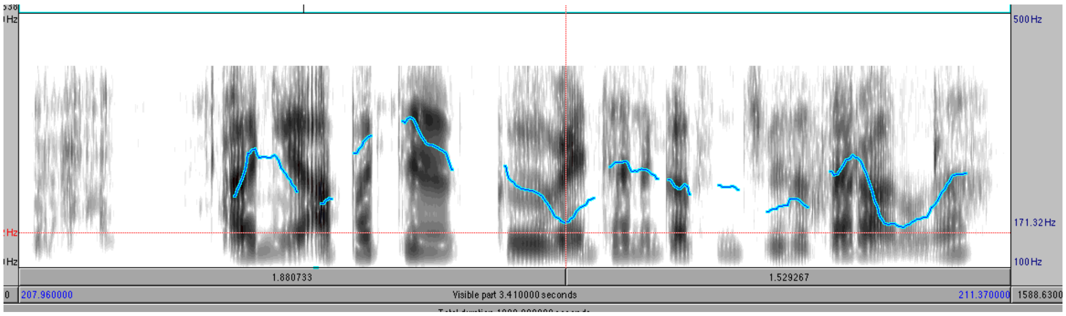
Utterances 200 and 415



These utterances also seem to associate Ls with sonorants but not plain segments; I don't have any particular explanation for these cases.

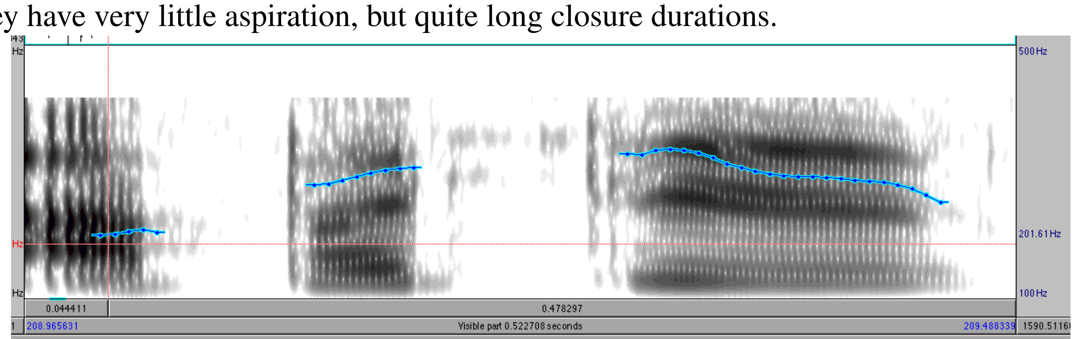
3.4.3 Speaker f6038

Utterance 207



The local minimum marked in the picture begins the English word *rent*. What's noteworthy here is the previous word, the determiner *ki*. This word is clearly external to the L-AP beginning with *rent*, but it's unclear if it forms an AP of its own. The pitch is level and it doesn't start low, as would be expected. This could be due to its short duration overall, or perhaps function words aren't always included in APs.

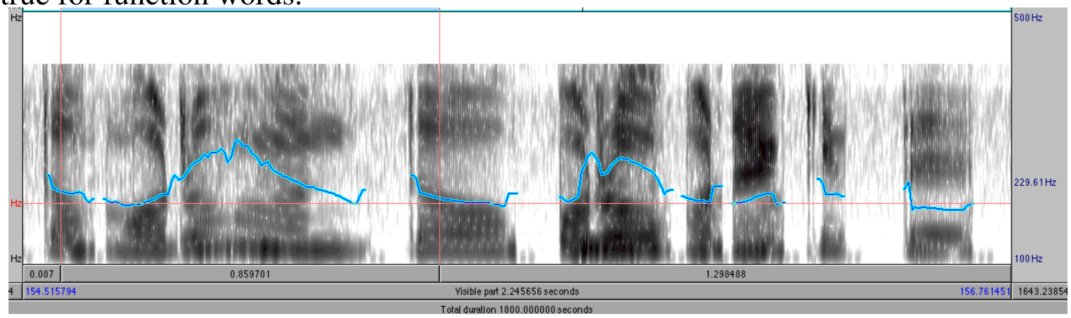
The word previous to *ki* in this example is *a-pha-thi*. I show a close up of this word below. It is a good example of two word-internal, underlying aspirated stops. You can see they have very little aspiration, but quite long closure durations.



3.4.4 Speaker f4546

Utterance 151

In this close up of a portion of Utterance 151, the two cursors each intersect an instance of the word *tto*. These words, which are particles and not content words, don't seem to form either an L-AP or an H-AP, but have mid-level pitch contours. This occasionally seems to be true for function words.



This speaker happens to be from Pusan, although I don't believe that's all that relevant to the question of function words being contained in APs. There are several speakers from Pusan in the corpus, and a cursory look at these speakers indicates that segment-induced tone is also present in this dialect. I didn't establish that the pattern of H-segments vs. L-segments is exactly the same, but there were clearly some instances of segment-induced H tones.

3.5 Conclusions

The examination of the sponataneous speech of speaker 6129 confirms most of what has already been observed about the phonetics of Korean obstruents and Accentual Phrases. L-APs and H-APs are frequently identifiable with full contours, and frequently contain multiple words. H-segments internal to the AP, whether underlying or derived, do not affect the pitch contour. The AP certainly appears to be the domain of plain stop alternations.

Jun's theory of tonal alignment however is contra-indicated by some examples from other speakers. Her theory is that the first two tones and the last two tones of an AP align with the first two syllables and the last two syllables, respectively. There are some examples of the second position H occurring after the second syllable. This is almost certainly a product of fast speech, and could be essentially an error, but it could mean that the alignment procedure is not actually syllable based, and only appears so in careful speech.

And there are clearly some oustanding issues, like to what degree up- or down-stepping occurs, and whether function words have the same status as content words in the formation of APs.

Finally, when it comes to the sound change in progress, it is difficult if not impossible to extract meaningful trend data from this corpus. It's clear however that plain stops and aspirated stops have quite a bit of overlap in their distributions even for speaker f6753 who was born in 1954. Even more generally, we can say that plain stops are much more like aspirated stops than unaspirated stops in spontaneous speech. It's unlikely that anyone could reliably distinguish plain from aspirated in this speech, if it weren't for tone and context of course.

Chapter 4

Change In Progress

The primary data collection of this dissertation presents a picture in apparent time of the properties of plain and aspirated stops in Seoul Korean, much like that reported in Silva (2006). VOT, as the primary correlate of aspiration, was examined for 20 speakers. Stop closure duration was also examined in the same data since this feature also contrasts the two manners, and so is relevant to a discussion of neutralization.

4.1 Experimental Design

Precision and efficiency were both of high priority in this experiment. As is always the case, there is a tradeoff between precision and the number of features that can be investigated. Robust and precise results require high repetition, but the more repititions for a variable the less variables could be examined, broadly speaking.

First, I decided to only examine one place of articulation. Research had shown that the three stop places patterned together; in other words, speakers that showed a receding contrast in aspiration showed it in all three places. I decided it was not a priority to demonstrate a change for all places, as long as I could demonstrate it for one place. Furthermore, research had also shown that there are consistent differences between places of articulation, in particular, velar stops tend to have longer VOTs than anterior places (both for Korean and cross-linguistically, see Chapters 2 and 5). Despite the phonological consistency of patterns, a phonetically precise investigation might be disserved by combining measures from different places. Therefore I chose a single minimal pair as the basis of this experiment: /pal/ and /p^{h}al/. Based on this pair and two other morphemes, six further words were constructed: /pal-mok/, /p^{h}al-mok/, /pal-til/, /p^{h}al-til/, /pal-mok-til/, and /p^{h}al-mok-til/. The morpheme /-til/ is a plural morpheme, which is not normally used, but nevertheless is grammatical. This yielded a set of 8 words which could be simply broken in half to examine the /p/ vs. /p^h/ contrast, or which could be broken down into finer groups, should the length of the word prove to be relevant. One additional advantage of this approach is that the number of tokens of /p/ and /p^h/ could be increased without simply repeating the same two words; the more variation in words, the less chance of causing speaker boredom and inattention.

The eight selected words were placed in the same carrier phrase: /nanin ne X-il tachota/, meaning "I hurt my X," or "my X hurts." Note that the target words gain an extra syllable in this phrase, due to the accusative suffix. This is important because it means the target words vary from 2 to 4 syllables total, and APs tend to be fully realized only when there are at least 4 syllables. For tonal measurements, it's possible that only those tokens, of length 4, are valid ones. In addition, the word /ne/ is an important part of the phrase. Without it, the juncture where the target phonemes lay would be between the subject and the object, probably the strongest phrase internal juncture. The addition of /ne/ puts the target phonemes in a more embedded position, which is potentially important for two reasons. First, domain initial strengthening effects might obscure or increase underlying contrast. This is a hypothesis on my part, but nevertheless I thought the contrast would be more representative in a prosodically weaker position. Second, it's more likely to observe true closure durations in this embedded position. At the stronger post-subject position, it's more likely that observed silence is partly due to a pause, and not solely due to stop closure. Nothing is perfect of course, and some speakers put unnatural pauses at word boundaries. Nevertheless, these were my rationales for constructing these sentences.

Table 4.1: Change in Progress Sentence Set

Korean	transliteration	gloss
발	pal	foot
발목	palmok	ankle
팔	p^hal	arm
팔목	p^halmok	wrist
발들	paltul	feet
발목들	palmoktul	ankles
팔들	p^haltul	arms
팔목들	p^halmoktul	wrists
carrier phrase: 나는 내 _______을 다쳤다		
"My _______ hurt(s)."		

Each speaker read a set of 80 sentences[1], 10 repetitions of the 8 sentences, in a random order. Each speaker read a random order[2] but not a unique order. Each random order created was read by roughly two or three people. It didn't seem necessary to have a unique random order for each speaker.

[1]Only 40 survived the recording session for speaker m03. Speaker f08 was recorded in two different sessions, yielding 160 tokens. She was one of my first speakers, and the first where I noticed the odd reversal of distributions described later. Since I was surprised at the result, I had her come back for a second session. Subsequently, I merged the data.

[2]Speaker f09 didn't read the sentences in random order, but read each sentence 10 times in a row before moving to the next sentence. She was the first speaker who recorded this set, and I didn't randomize it because I hadn't decided to use it widely yet.

4.2 Results

4.2.1 VOT

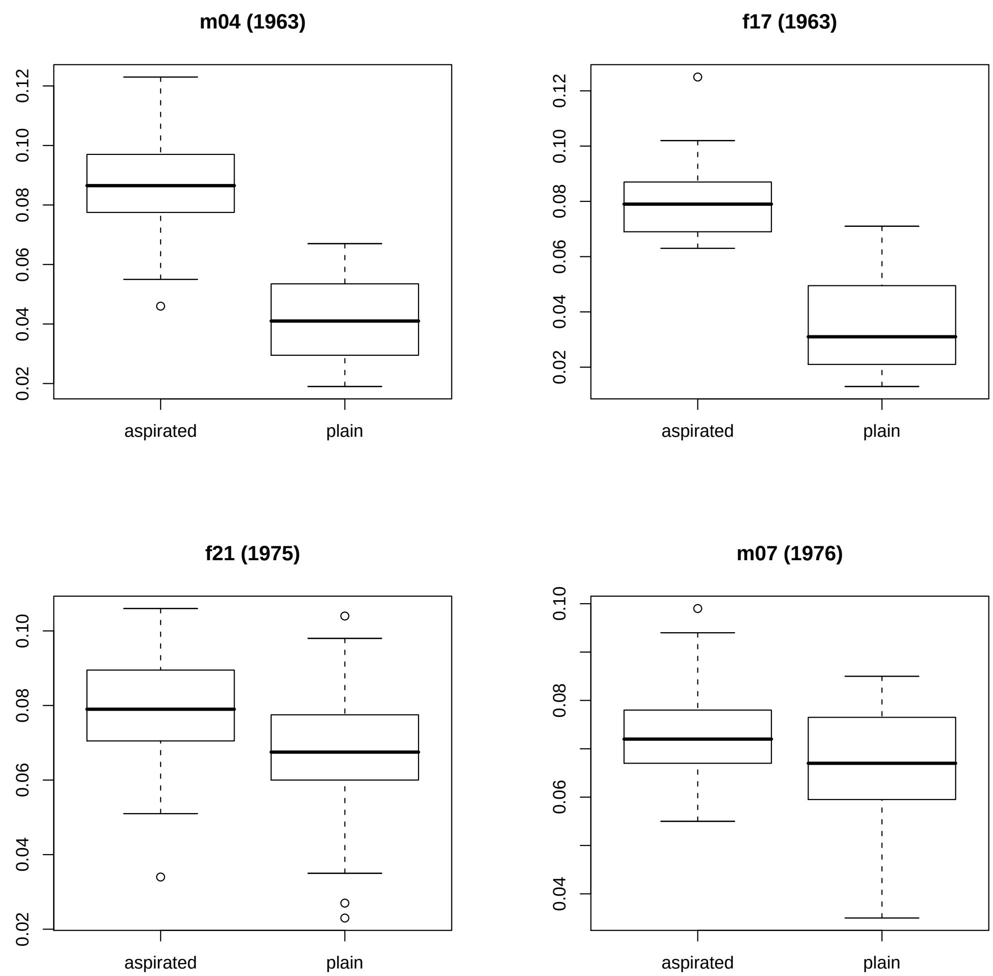
The following table shows the VOT results for /pʰ/ vs. /p/ for all speakers, in apparent time chronological order. Please note the following two conventions: all data in this chapter is presented in the order aspirated first, plain second; aspirated and plain are represented as h+ and h- in the boxplots, due to the notation I used in the R statistical analysis program, the mnemonic being "plus or minus aspiration."

Table 4.2: Aggregate VOT data

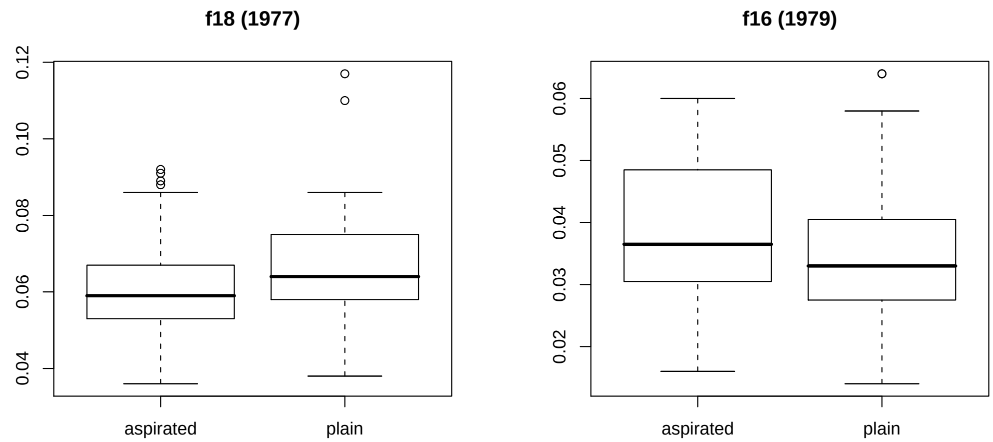
speaker	YOB	p^h	p	t-test, p	wilcox test, p
f20	1955	60	34	$<< .0001$	$<< .0001$
m04	1963	87	41	$<< .0001$	$<< .0001$
f17	1963	85	35	$<< .0001$	$<< .0001$
f22	1971	53	39	$< .0001$	$< .0001$
f23	1971	59	43	$< .0001$	$< .0001$
m03	1972	48	32	$< .0001$	$< .001$
f09	1972	64	50	$< .01$	$< .01$
f21	1975	78	68	$< .01$	$< .01$
m07	1976	73	67	$< .01$	$< .05$
f18	1977	61	67	$> .05$	$> .05$
f13	1979	38	24	$<< .0001$	$<< .0001$
f16	1979	38	35	$> .1$	$> .1$
f05	1980	70	81	$< .0001$	$< .0001$
f14	1980	28	19	$< .0001$	$< .0001$
f08	1983	76	80	$< .01$	$< .01$
m05	1983	53	34	$<< .0001$	$<< .0001$
f01	1985	42	44	$> .1$	$> .1$
f15	1986	42	65	$<< .0001$	$<< .0001$
f19	1986	47	47	$> .1$	$> .1$
m06	1987	37	51	$< .001$	$< .0001$

I'll illustrate the variation in VOT with distribution plots from a few individuals. Speakers m04 and f17, both born in 1963, show the traditional pattern with a clear separation between the two manners. Speakers f21 and m07, born in 1975 and 1976, have distributions that are not clearly distinct, but that are still slightly significantly different.

Figure 4.1: VOT distributions for m04 and f17

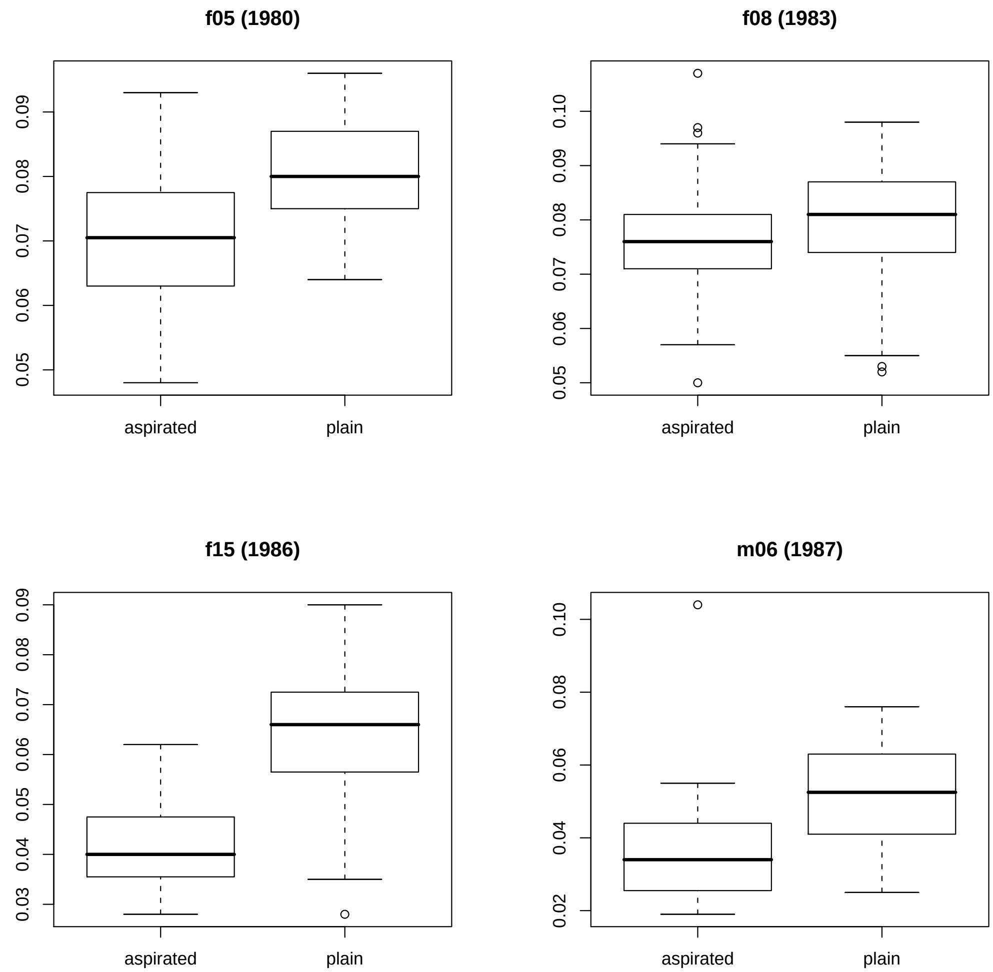


Speakers f18 and f16, born in 1977 and 1979, have distributions which are not significantly different.



I expected to see much more of this pattern, but technically these are the only two true natives of Seoul with equivalent distributions. Some young speakers from Seoul, 4 in my data set, have significantly different distributions where the plain stop VOT is *longer* than the aspirated stop VOT (this was also reported in Silva (2006)). The difference in means is very small for speaker f08 for whom there is double the data, but the difference in means for speaker f15 is 23 ms, which is not small. I believe these speakers also represent a merger in aspiration, despite the differing distributions; I discuss this further in the conclusory chapter. The distributions for these four speakers are shown on the next page.

Figure 4.2: VOT, f05

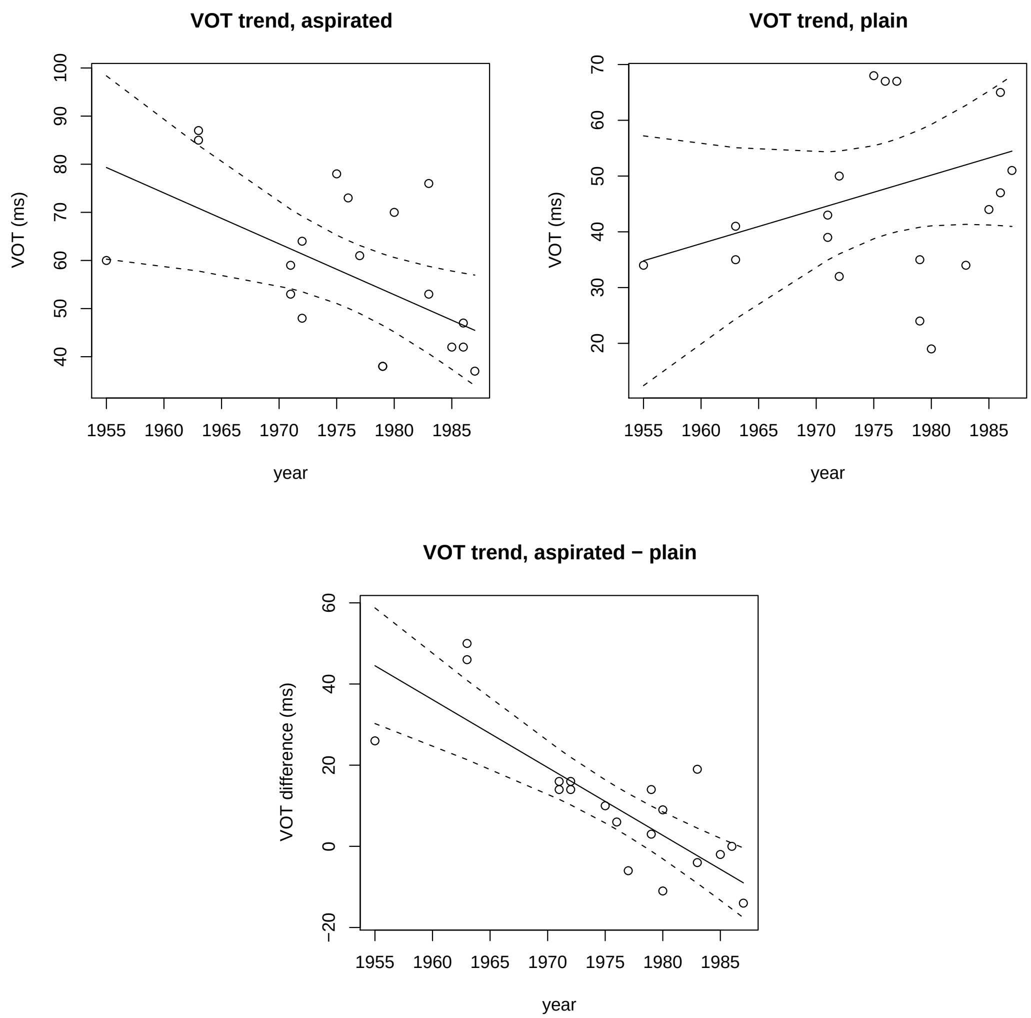


Speakers f01 and f19 have equivalent distributions (no significant difference), but are not true Seoul natives. Speaker f01 lived in the US from birth until age 7, then moved to Seoul. Speaker f19 is from Kyungsang, and went to Seoul for college. However, her father is from Seoul, and other Koreans are surprised to hear that she's not from Seoul[3]. It would be one thing if these speakers maintained the contrast, but since they do not, that suggests they're participating in the Seoul pattern.

[3]I witnessed the surprise of her two friends when she confessed she was from Kyungsang. Her friends think she sounds just like any Seoul native.

The following scatterplots and linear regression lines illustrate the VOT trends: aspirated VOT is decreasing and plain VOT is increasing, which is expressed in the third plot as the decreasing trend of the VOT difference.

Figure 4.3: VOT, f05



4.2.2 Closure Duration

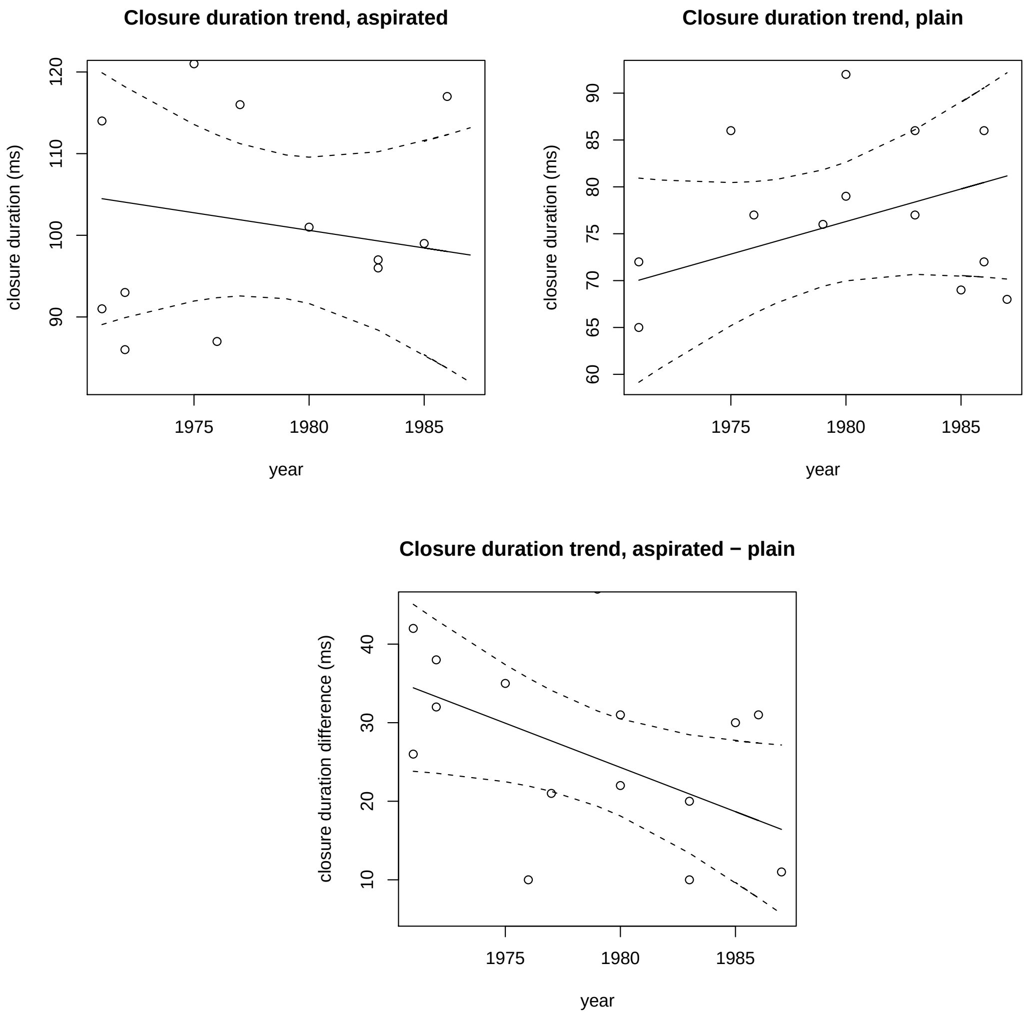
Measuring closure duration was slightly problematic since sometimes people pause slighly immediately before the target word, and if they do, the measurement isn't reflective of inherent duration. Three speakers were excluded entirely since they consistently read the sentences with pauses. Other speakers at least occasionally did so, probably out of momentary uncertainty, while still others were quite fluent in their pronunciation. However, almost all speakers had a significantly different closure duration pattern, with aspirated stops having longer closure durations.

Table 4.3: Aggregate closure duration data

speaker	YOB	p^h	p	t-test, p	wilcox test, p
f20	1955	109	63	$<< .0001$	$<< .0001$
f22	1971	114	72	$<< .0001$	$<< .0001$
f23	1971	91	65	$<< .0001$	$<< .0001$
m03	1972	86	54	$<< .0001$	$< .0001$
f09	1972	93	55	$<< .0001$	$<< .0001$
f21	1975	121	86	$<< .0001$	$<< .0001$
m07	1976	87	77	$< .001$	$< .01$
f18	1977	116	95	$< .05$	$< .0001$
f13	1979	123	76	$<< .0001$	$<< .0001$
f05	1980	101	79	$<< .0001$	$<< .0001$
f14	1980	123	92	$<< .0001$	$<< .0001$
f08	1983	97	77	$<< .0001$	$<< .0001$
m05	1983	96	86	$> .1$	$< .05$
f01	1985	74	72	$> .5$	$> .5$
f15	1986	99	69	$<< .0001$	$<< .0001$
f19	1986	117	86	$<< .0001$	$<< .0001$
m06	1987	79	68	$< .0001$	$< .0001$

Trend plots are also given on the following page, which match the trends seen for VOT. Since the two speakers born in 1963 were excluded for the reason mentioned above, I decided to also exclude the speaker born in 1955, for fear that a single speaker so far from the others in terms of age might inappropriately skew the results. It turns out that including that speaker barely changes the trend line, and actually makes it more conservative.

Figure 4.4: VOT, f05

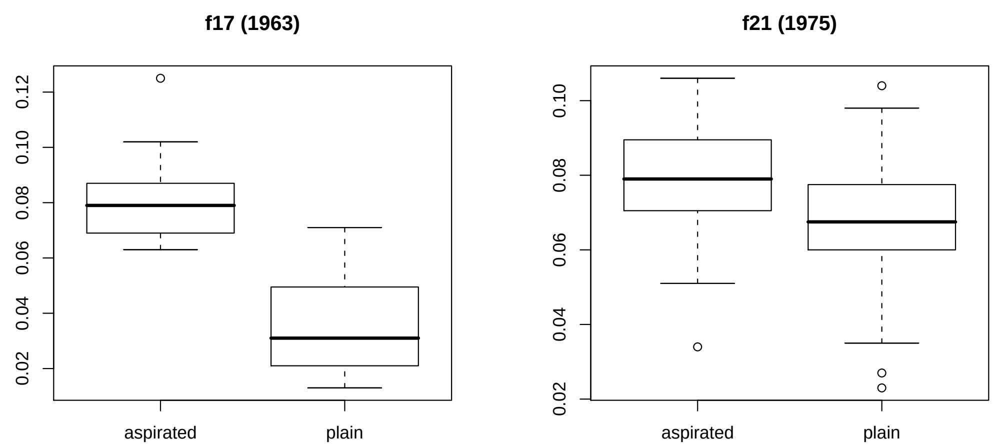


4.2.3 Summary

These speakers show a trend towards reducing, eliminating, and even reversing the traditional aspiration contrast between aspirated and plain stops. This reversal of contrast is not simply a statement about the trend line falling below zero: 4 speakers actually have significantly greater VOT means for their plain stops, as much as 23 ms greater for one speaker. 4 additional speakers have no significant difference between the two means. Closure duration shows a similar trend of a narrowing difference between the two manners, but no more than two speakers show a complete neutralization in these terms (but see Chapter 5). All other speakers shown have significantly greater duration means for their aspirated stops. For Speaker m05, the p-value for his data falls below the .05 significance level for the Wilcox test but not the T-test, but his aspirated VOT mean is signficantly greater than his plain VOT mean. Speaker f01 on the other hand has clearly non-distinct closure duration and VOT distributions, perhaps due to not quite being a true Seoul native.

One more observations can be made about the change underway that is not evident from the trend plots. Consider the following two plots that show the VOT distributions for speakers f17 and f21, separated in age by 12 years.

Figure 4.5: VOT plot, speakers f17 (1963) and f21 (1975)



The phonetic correlates of phonemes are not means, they are distributions. When viewing the change in these terms, the two maners are undergoing the same change: they are expanding. While older speakers have primarily distinct distributions, young speakers have expanded distributions with substantial overlap. This expansion is considered further in following chapters.

Chapter 5

Phonetic Neutralization

While a cursory look at the change in progress in Seoul Korean suggests that the contrast between plain and aspirated stops has been neutralized in AP-initial position, the finer view presented in Chapter 4 calls that conclusion into question: for most speakers the VOT and closure duration distributions were significantly different. There are also two odd features about the trends of these acoustic measures. First, there is a trend towards a reversal of the VOT distributions, where plain stops have longer VOTs than aspirated stops. Second, the closure duration distributions seem to lag behind the VOT distributions in apparent time: for all but one or two speakers the aspirated stops still have longer closure durations than the plain stops.

The following sections support the hypothesis that some of the speakers in this study do show a true neutralization, and that the above patterns are the automatic phonetic consequences of the coarticulation of tone.

5.1 Measuring VOT

5.1.1 The Assumption

Measuring the phonological feature aspiration is not as straightforward as it at first seems. VOT is the primary acoustic correlate of aspiration, but modes of phonation cannot be ignored: a period of breathy phonation could serve as the cue to aspiration. Hypothetically a Korean speaker could maintain the contrast between plain and aspirated stops by having breathy voicing after the aspirated stops such that VOT was identical to plain stops. This clearly isn't the norm, but should be possible. Speaker f14 has abnormally short VOT values, and on a second look at her recordings, it was clear that she did have a lot of breathy voicing compared to other speakers. It's likely that making her data comparable to the other speakers requires taking that breathy voicing into account in the measure of VOT.

The change in progress analysis was based on a methodological hypothesis that measuring VOT based on periodic motion in the waveform was more accurate than using the spectrogram. The potential tradeoff was that breathy voicing or random low amplitude periodic motion would be unduly recognized. I made the assumption that transitional motion in the waveform, between the regions of no phonation and modal phonation, would be consistent within each speaker across plain and aspirated manners. Both these manners involve abducted vocal folds that are brought together to produce modal phonation. As the vocal folds begin to reach each other, it's common to have some irregular, non-modal phonation, but to the extent that this is an idiosyncratic effect of the speaker's vocal folds, it should average out to be the same for the two stop manners. Therefore, this approach to measuring VOT would accurately capture any contrast. That was the assumption. The hypothetical speaker mentioned in the previous paragraph would present a problem to this approach since the breathy voicing was a cue inherent to one manner but not the other, but again, this is not the norm. In addition, speakers with relatively long VOTs by my metric are unlikely to employ breathy voice phonologically, because that would mean three distinct modes of phonation from stop to vowel: no phonation, breathy phonation, and modal phonation. It's much more likely that breathy phonation present in the transition is an unintended phonetic consequence.

5.1.2 The Problem

However, the reversal of VOT contrast called to my attention a problem with my assumption of basic equality between the articulations of the two manners. When faced with an apparent reversal of VOT contrast for some speakers, my conclusion was that the laryngeal settings for the two stop manners were not parallel, due to tone, and that this affected the VOT measurements. In hindsight this possibility should have been evident from the beginning: the high and low pitch following the different manners is immediate, so the larynx must already be set for those pitch values during stop articulation. Unfortunately the superficially independent functions of the larynx are not completely independent in reality. While the raising of pitch via movement of the cricoid cartilage is an independent articulation from glottal abduction, it changes the vocal folds which might fundamentally change all aspects of phonation, not just pitch. As Mark Liberman pointed out to me, a stretching of the vocal folds due to cricoid movement necessarily brings them closer together because the end points of the vocal folds are already maximally separated. This effect, possibly, is the source of the VOT reversal: if speakers initiate adduction at the same time for the two stop manners, the aspirated stops should begin phonation sooner because the glottis is actually narrower overall.

In short, my hypothesis is that the articulation of tone can have an automatic phonetic effect on VOT. If this is true, the effect could skew the distributions in statistically signficant ways while being unintentional and non-contrastive.

5.2 Precedent for Automatic Laryngeal Interactions

5.2.1 Shanghainese

Previous Study

Shen et al. (1987) report data for Shanghainese that is strikingly similar to the situation in Korean. In Shanghainese there is a three way contrast between voiced, voiceless unaspirated, and voiceless aspirated stops. At least, this is the historical contrast. The so-called voiced series is de-voiced in initial position but voiced word-internally (inter-vocalically). This voicing neutralization followed a split in the tonemes following the classic tonogenesis pathway: the historically voiceless stops can only begin words with one of the three high register tones, and the historically voiced stops can only begin words with one of the two low register tones. Therefore descriptively in both Korean and Shanghainese we have very analogous situations: a word-intial voiceless series that is voiced word-internally, but word-initially contrasts with another voiceless series by means of a lower tone (and potentially some other laryngeal feature).

There is further similarity in the fine phonetic details, as can be seen in the table below. In the tone column, the number is the authors' means of identifying the tones, and is only relevant when refering back to the article. The H/L symbol identifies the register, as tones 1, 2, and 4, begin high as compared to tones 3 and 5. The data come from all three places of articulation, labial, alveolar, and velar. Also recall that the L values represent stops that were historically voiced initially and that are still voiced inter-vocalically.

Table 5.1:

tone		closure duration				VOT			
		mean	SD	min	max	mean	SD	min	max
1	H	157.80	15.26	128.00	191.50	16.27	11.94	5.90	48.00
2	H	150.60	15.39	116.60	178.50	19.45	9.25	7.20	42.30
4	H	156.26	14.15	132.40	179.70	15.52	9.02	7.90	34.00
3	L	126.01	19.92	82.80	166.10	24.33	10.70	11.50	45.70
5	L	125.98	13.65	99.60	149.80	21.43	12.24	6.90	59.40

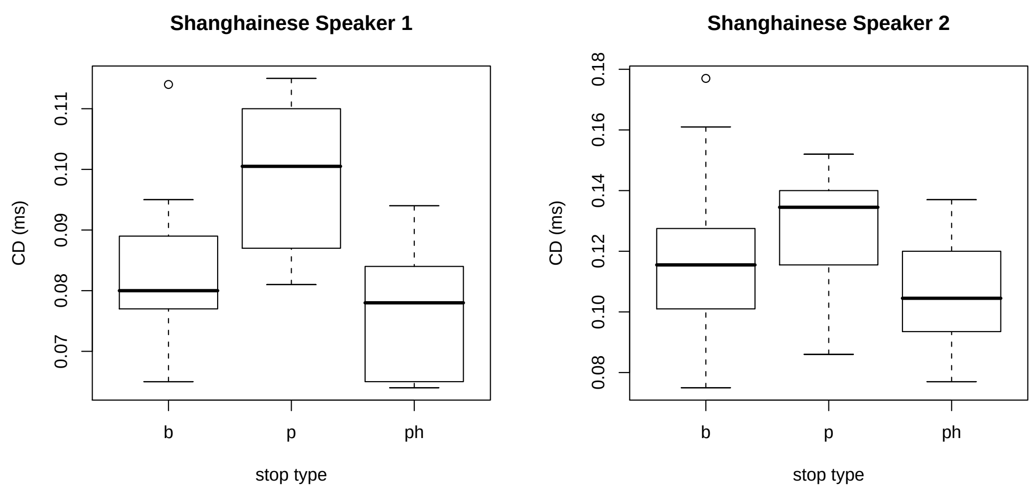
Table 5.2: N = 30 for all means.

While all the stops represented above are voiceless, the ones associated with higher tones have longer closure durations and shorter VOTs. These are exactly the correlations I found for the younger speakers of Seoul Korean. Could Shanghainese and Seoul Korean be in the midst of similar changes? Where a tonal contrast replaces a voicing contrast in Shanghainese and an aspiration contrast in Korean? I think the more important question is, do superficially orthogonal phonetic measures have inherent, measurable dependencies?

The Shanghainese data supports my hypothesis that high tone lowers VOT through coarticulation. But what about the closure duration? Before seeing this data, I believed that the closure durations in my Korean data indicated that neutralization hadn't taken hold in the community since almost every speaker has longer durations for their aspirated stops. After seeing this data, the question has become, can the coarticulation of tone effect closure durations? Just as it seems that tone can effect VOT distributions, perhaps it can effect closure durations as well. This would immediately solve one problem with my Korean data, the fact that the closure duration trend seems to lag behind the VOT trend. If high tone causes longer closures, as well as shorter VOTs, this would suggest that several of my Korean speakers actually have neutralized the plain vs. aspirated distinction in initial position, and that the two phonetic trends are proceeding in concert.

Reproduction

I tried a limited reproduction of the above study, including only the labial stops and only two tones, one high and one low (Tones 1 and 3 in the authors' description). I also added the labial aspirated stop, for the high tone only. My informant couldn't think of a word with the aspirated labial stop and the low tone, and not knowing whether that was a real constraint in the language or not, I decided to proceed without that combination. Therefore the following plots compare /p/ (high), /b/ (low), and /p^h/ (high), for two speakers.



The data for Speaker 1 includes ten tokens of each type. For Speaker 2 there are 40 tokens of each type, 20 each from two different syntactic positions. I attempted to replicate the previous research by eliciting word-initial and word-internal positions (for the second speaker only). The authors claim that word-internally the voiced series is still voiced (unlike word-initially), but in my data they were voiceless in this position also. I simply may not have replicated the experiment correctly. Regardless, I merged the data from both positions.

There is a consistent pattern, on the whole and within different subtrials, of the /p/ series having longer closures than both of the other series. This is consistent with the previous study, although the differences are much smaller in my data. While the previous study compared VOT values, in my data, the VOTs for the unaspirated stops were too short to be meaningfully compared. It's possible that differences in elicitation resulted in prosodically strengthened values in the previous study. It's also possible that the inclusion of velars in the previous study skewed the results, although we would only expect this for VOT, and not closure duration (see discussion below on velars).

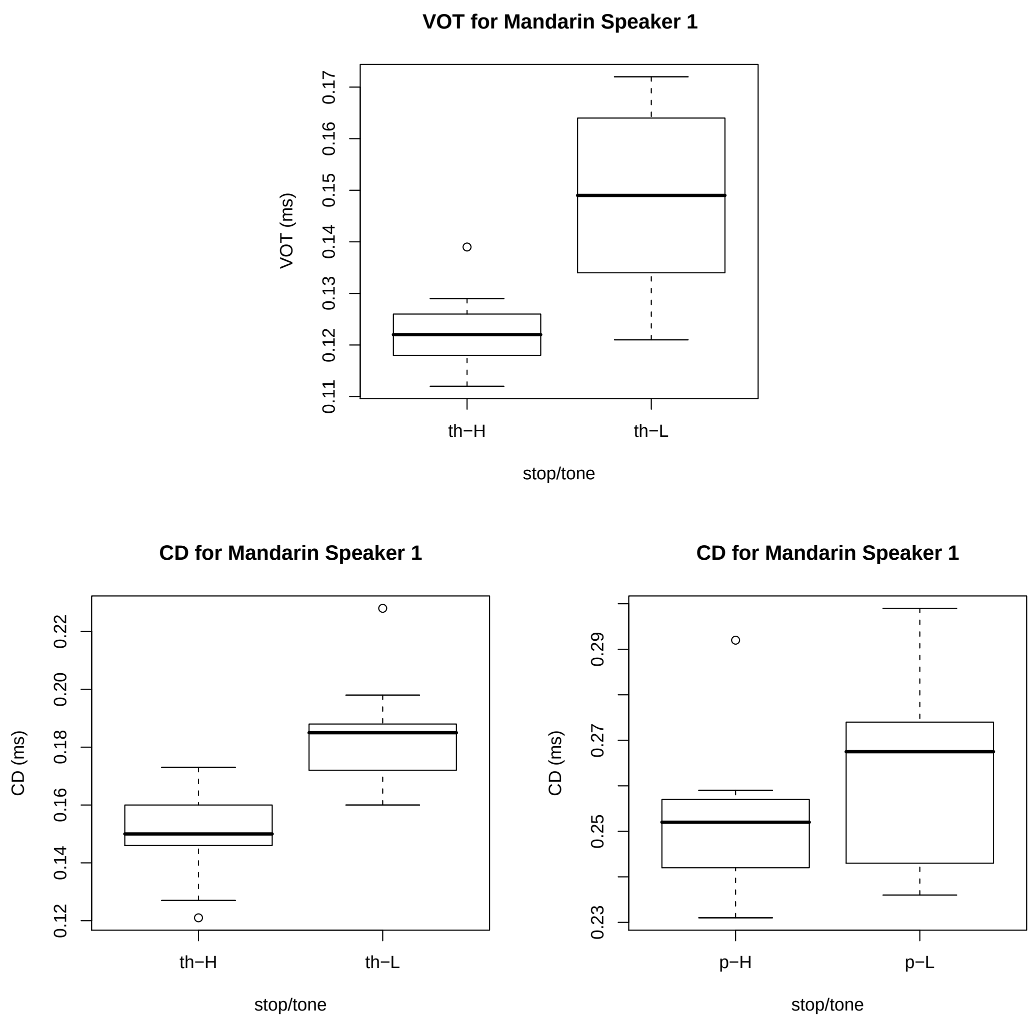
It's important to note that while the /p/ and /b/ values show a positive correlation between pitch and closure duration, the /p^h/ values have high tone but have closure durations more similar to /b/ than /p/. I'll return to this issue below.

5.2.2 Mandarin

In Korean, the question is whether tone can have an automatic effect on either VOT or closure duration. To explore this I conducted some small trials on Mandarin, which has lexical tone and two series of stops, voiceless aspirated and voiceless unaspirated. The results from two male speakers are shown below.

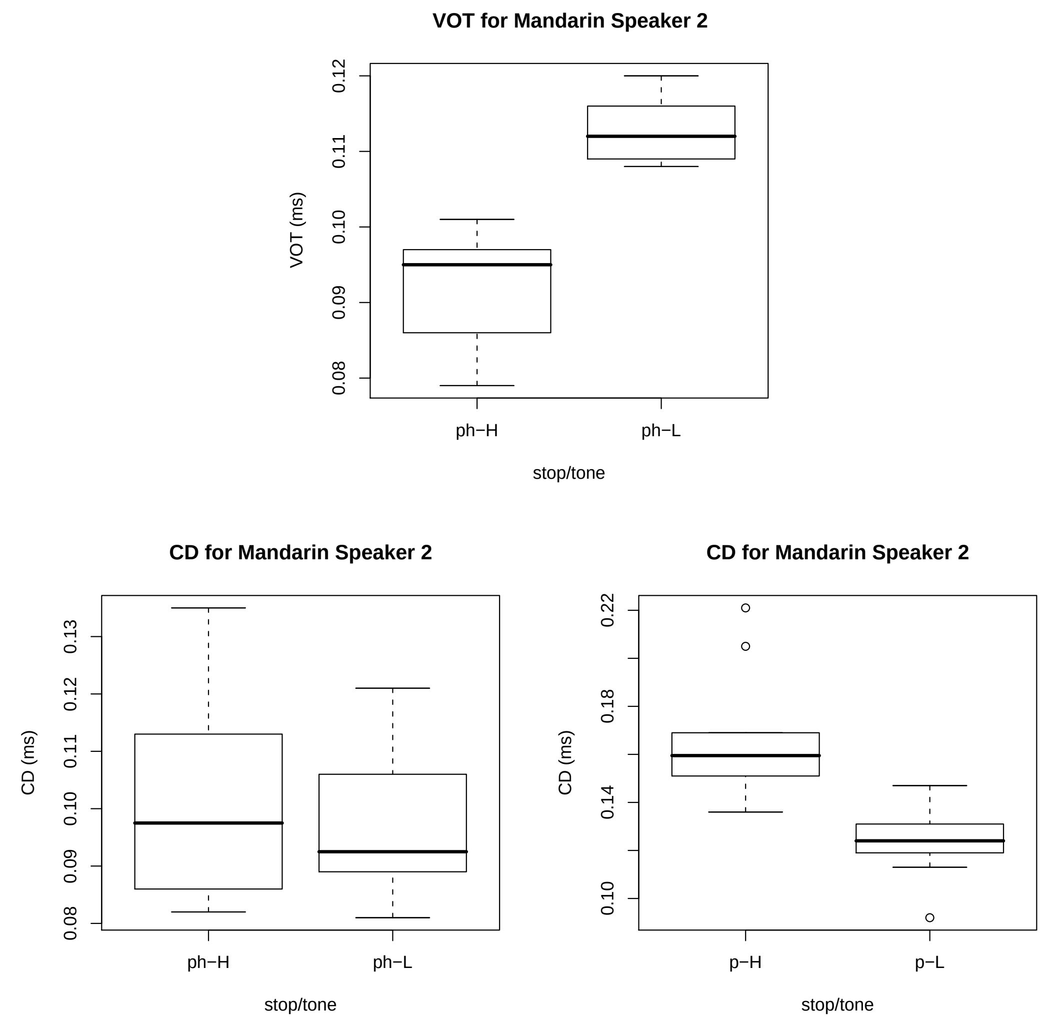
Mandarin is typically described as having four contrasting tones with specific patterns over a 5 level pitch scale. Without going into the details of each tonal pattern, and only considering the beginning of each tonal pattern to be relevant here, it's roughly true that Tones 1 and 4 are high, and Tones 2 and 3 are low. The data below for the first speaker contrasts Tones 1 and 3 as representatives of high and low tone, respectively. The stops from ten tokens each of /t^ha-1/, /t^ha-3/, /pa-1/, and /pa-3/ are compared (the mismatch in place of articulation is a flaw, not an advantage).

Figure 5.1: Data for Mandarin Speaker 1



For the second speaker's data, shown below, Tones 4 and 2 are used as representatives of high and low tone, and the stops came from ten tokens each of the following: /p^ha-4 p^ha-2 pa-4 pa-2/.

Figure 5.2: Data for Mandarin Speaker 2



Both speakers show significantly longer VOTs for low tone. As for closure duration, Speaker 1 has significantly different distributions for the aspirated series, but not the unaspirated series. The reverse is true for Speaker 2. Furthermore, the effect of low tone is a lengthening one for Speaker 1 but a shortening one for Speaker 2.

These trials are preliminary and far from conclusive, since the token count is not high, and the experimental design was not as controlled as it could have been. However they do support the idea that there are interactions between pitch and other phonetic features. The inverse correlation between pitch and VOT seen in the Korean data is also seen here. And there is evidence that pitch and closure duration sometimes interact, although the interaction may be speaker dependent.

However, I don't mean to suggest that the interaction is *arbitrary*, in the sense of arbitrary, learned phonetic implementation rules. For example, it's known that voiced stops often involve lowering of the larynx and/or pharyngeal exansion. Both these articulations expand the oral cavity, increasing the glottal pressure difference, which allows for longer voicing during closure. Just as increased pressure in the oral cavity seems to affect the properties of velar stops, the decreased pressure of voiced stops might lengthen their closures. If this was true, and the oral cavity expanding articulations were speaker dependent, there would be a speaker dependent interaction between voicing and closure duration that was still based on automatic phonetic principles. This is an important distinction. Rather than phonetic differences that are learned and therefore *could* result in a phonological contrast, purely automatic differences that can *not* be learned may arise in a speaker dependent fashion due to the freedom of the system. Consider the discussion in Nearey (1980) on the idiosyncracies of vowel articulation.

5.2.3 Further Interactions

While the above data is not conclusive, it did change my hypotheses about what was going on in Korean. While I don't have a specific theory as to how pitch could affect closure duration, there are well-known laryngeal-closure interactions. Firstly, velar stops usually have longer VOTs than more anterior stops. Not only is this pattern clear from the Korean phonetic literature alone, but Cho and Ladefoged (1999) find this to be the predominant pattern in their survey of 18 languages. They discuss several factors that could contribute to this pattern, including a theory put forth in Maddieson (1997). Velar closures tend to be shorter than those of more anterior places, presumably due to the smaller cavity behind the closure that leads to a more rapid build-up of air pressure. If the laryngeal gestures of voicelessness and aspiration are inherently independent of place of stop articulation, the shorter closure of velar stops will result in longer VOTs. Therefore gestures that are identical at some level of abstraction can lead to automatic phonetic differences for physical reasons.

The well-known study by Lisker (1957) showed that in English, intervocalic voiceless stops are longer than intervocalic voiced stops[1]. However, it's important to note that the test stops of that study began word-internal unstressed syllables, for example *rapid* vs. *rabid*. Cole et al. (2007) examined the properties of stops in a radio corpus, with particular focus on the role that phrasal accent might play. In that study, the test consonants were always word-initial, but the syllables that they began may or may not have had a phrasal accent, and when unaccented may or may not have had stress. Unlike Lisker (1957), in their data voiced stops have longer closures, consistently. A small sampling of my own speech supports these two patterns, with the addition of word-internal pre-stress position, as in *repel* vs. *rebel*, patterning like word-initial position: voiced stop closures were slightly longer. If you compare the plots for the Mandarin speakers above, you can see that the closure durations of the unaspirated stops are much longer than those of the aspirated stops. This pattern also appeared in one of the subtrials for Shanghainese Speaker 2, although it's not evident in the merged data above. It's possible that there is a reverse correlation between closure duration and VOT across stop types in initial position, just as there is across place of articulation. In other words, it seems that, in terms of closure alone, voiced stops are

[1]Jessen (1998, pp.62-64) reports the same for German, that word-internal intervocalic voiceless stops are longer than voiced stops; on the other hand, word-initial intervocalic stops don't show a consistent pattern for closure duration

longer than voiceless stops, which in turn are longer than aspirated stops. If there's a tendency for word-initial stops to be the same total length, then this pattern could aslo have a compensatory explanation.

5.2.4 The Problem with Korean

So, the relationships between place of articulation, phonation type, closure duration, and VOT seem to involve automatic phonetic effects. It also seems likely that there's an automatic phonetic interaction between tone and VOT (based on the Korean and Mandarin data). Therefore some sort of automatic interaction between tone and closure duration is not unlikely.

But Korean doesn't fit completely into the above patterns when everything is brought together. The positive correlation between pitch and closure duration and the negative correlation between pitch and VOT does fit with Mandarin and Shanghainese. But this is only sensible if we consider both the plain and aspirated series in Korean to be aspirated (which is true at least phonetically). In Mandarin, Shanghainese, and English (word-initially), aspirated stops have *shorter* closure durations, but in Korean the more aspirated series has *longer* closure durations. If we believe that the Korean plain series was originally unaspirated, the cross-linguistic facts would lead to the conclusion that in the past the plain series had longer closure durations than the aspirated stops, which is contrary to the current trend in apparent time. It's almost certainly not true that the closure duration distributions converged, passed each other, then reversed direction and converged again. Therefore, there is still a missing piece to this puzzle. It's likely that automatic phonetic effects aren't completely explanatory for the observed patterns, and that we have to allow for arbitrary, learnable differences.

5.3 Neutralization and Change

5.3.1 Just Noticeable Differences

Much of the argumentation in this dissertation involves what is and what is not perceived as different. In phonetics, research has been done on the Just Noticeable Difference (JND) for different acoustic parameters. For example, the JND for pure tones is around 1 Hz (for tones up to 1 KHz), and for formant-like bands of energy, around 10 to 20 Hz (Stevens, 1998). But if these are phonetic measures, what would a phonological measure be? Phonemes represent distributions which are often overlapping; if a learner was exposed to two phonemes whose difference of means was close to the level of a JND, it's a near-certainty that they would merge the phonemes because the distributions would overlap so much. If there is a JND for an acoustic measure, there must also be a Phonemic-JND (P-JND) for that measure, a minimal acoustic difference of distributions that could maintain the contrast. Whether accurate P-JND values for real human-beings could be empirically measured, I don't know, but the values must exist. For two distributions with known shapes (e.g. normal distributions) and known standard deviations, and for a given number of values extracted from these distributions at random, what is the minimal difference of means that would allow a learner to recognize that there are two distributions rather than just one? That value would be the P-JND.

Consider the puzzling cases of near merger, where people cannot perceive a difference that they can produce (Labov, 1994). Mergers cannot be undone, but near mergers can be, obviously, because they are not actually mergers. In the case of a reversal of a near merger, where the phonemes diverge, distance between the phonemes must not have fallen below the P-JND, because the community was able to propagate and extend the distinction. In a scenario where speakers have a production/perception mismatch, but the community then proceeds to reverse the near merger, we may presume that a particular person's conscious perceptions are not as accurate as a learner's or a linguist's. But is it possible for a near merger to fall below the level of the P-JND? Leave aside for the moment the question of how we verify the near merger, whether through a linguist's perceptions or through instrumental measure. Is it possible for a speaker to have such a narrow convergence of phonemes that the distinction would be opaque to learners and would not be propagated? My hypothesis is yes, it is possible. Keep in mind that the *means* would most likely be perceptibly different, as JNDs are quite small. A P-JND is a statement about distributions and acquisition.

So, how could a speaker produce a distinction, *learn* a distinction, that is so close that it can't be passed on to learners? The scenario sounds contradictory because the speaker in question had to acquire the distinction to begin with. I think there are at least three answers to this, the first of which is simply learner error, an incorrect estimate of means that puts them closer together. Second, random variation inherent in the learner's system, and the production apparatus may establish two different production means that don't quite match what was perceived. However, both of these factors should create divergence just as often as they create convergence, so probably wouldn't explain a community-wide pattern. I believe the third possibility is the most important: post-acquisition change within the speaker. This goes straight to the issue of the *incrementation problem* in linguistic change.

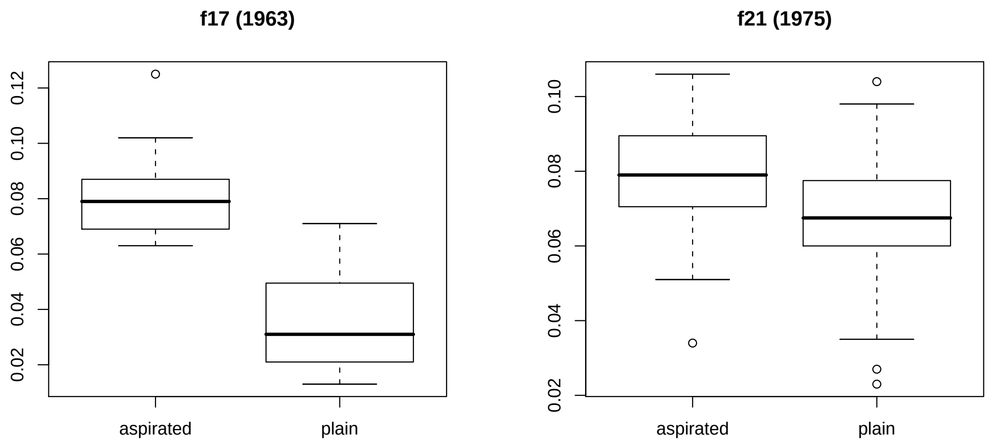
5.3.2 The Incrementation of Linguistic Change

Another puzzling aspect of linguistic change is its distribution over generations. If change is due to acquisition error, why doesn't it stop after one generation? Or go backwards? It seems that generation after generation keeps making the same mistake in the same direction, at least in some cases, like the movement of a vowel through the vowel space. If change is due to peer-pressure and the expression of social identity, why does the sociolinguistic variable continue to diverge rather than stabilize? The answer may lie in a speaker's ability to change after acquisition. This begs the previous questions. If change is due to learner error, why would the learner continue to err later in life rather than correct their mistake? If change is driven by social identity, why would a learner continue to change rather than stabilize at the desired value of the sociolinguistic variable? If the incrementation of a change over generations is due to post-acquisition change, then in each case there must be a specific force that drives that change within the individual, whether it's an internal or external force. In the case of a change that is leading towards neutralization, that force could drive an individual's distinction into a situation where the difference in means falls below the P-JND threshold.

In other words, a solution to the incrementation problem could be a dynamic model of the individual: rather than simple acquiring a phonetic grammar that remains fixed, there might be change during the lifespan. Such a model would also allow the situation where a learner produces a difference so narrow that it can't be perceived: the learner originally produced a wider difference, which later narrowed below the level of the P-JND.

Case in point, Korean aspiration. Consider the VOT plots for speakers f17 and f21 in Figure 4.5, repeated from Chapter 4, where I point out that the VOT distributions of both manners are expanding in apparent time such that they overlap more for younger speakers. For speaker f21, the VOT means of 78 and 68 are significantly different. The horizontal lines represent quartiles, so over 75% of each distribution is overlapping. Furthermore, the JND for duration is around 10 ms (Stevens, 1998). And, this is laboratory speech, not natural speech. It seems extremely unlikely that a learner, based on f21's input alone, could identify two different distributions. If this is true, this would imply that f21 could not have acquired her distinction based on input like her own output. Of course, we basically know that she was not exposed to input like this, simply because there is a change in progress and older speakers have more distinct distributions, like speaker f17. But beyond this, I'm saying that she *couldn't* have faithfully replicated this pattern directly, because it's not a perceptible difference. It's possible that speakers like f21 acquired a contrast that was decreased by post-acquisition change below the level of the P-JND.

Figure 5.3: VOT plot, speakers f17 (1963) and f21 (1975)



5.4 Phonetic Forces

In previous sections I elaborated on incrementation and the Seoul sound change by suggesting speakers are subject to a force that propagates the change even after acquisition. But what force?

To combat the effect of least effort and commuincate effectively, other internal forces must maintain the positions of phonemes in perceptual space. Whatever the mental representation for aspiration is precisely, some force must maintain a boundary between the plain and aspirated distributions to keep mispronunciations in check. The acquisition of a language, rather than a fixing of the linguistic system, could be viewed as the stablization of a still dynamic linguistic system. I believe that in Seoul Korean the perceptual salience of tone has taken attention away from aspiration and therefore relaxed the "boundary" force between the plain and aspirated distributions. The principle of least effort would then naturally expand the distributions, even after acquisition. Because the aspiration space is effectively closed, with a definite minimum (zero), and essentially a maximum as well (due to the following vowel), the expansion will also move the centers of the distributions together.

In short, if there is no need, no pressure, to maintain the aspiration difference, the system will tend to degrade and not preserve that difference. On the other hand, why now and why only Seoul? Presumably there was a trigger, for which I don't have a hypothesis.

Chapter 6

Perception Experiment

To support the analysis of change in progress, a perception task was performed by most speakers discussed in that chapter. The goal was to test speakers' sensitivity to both VOT and pitch in their identification of the phonemes /p/ and /p^h/, by taking natural tokens, modifying their properties, and presenting them to the speakers. The hypothesis was that pitch would be a more important cue than VOT, and that there would be an age related bias, younger speakers showing more reliance on pitch than older speakers.

6.1 Experimental Design

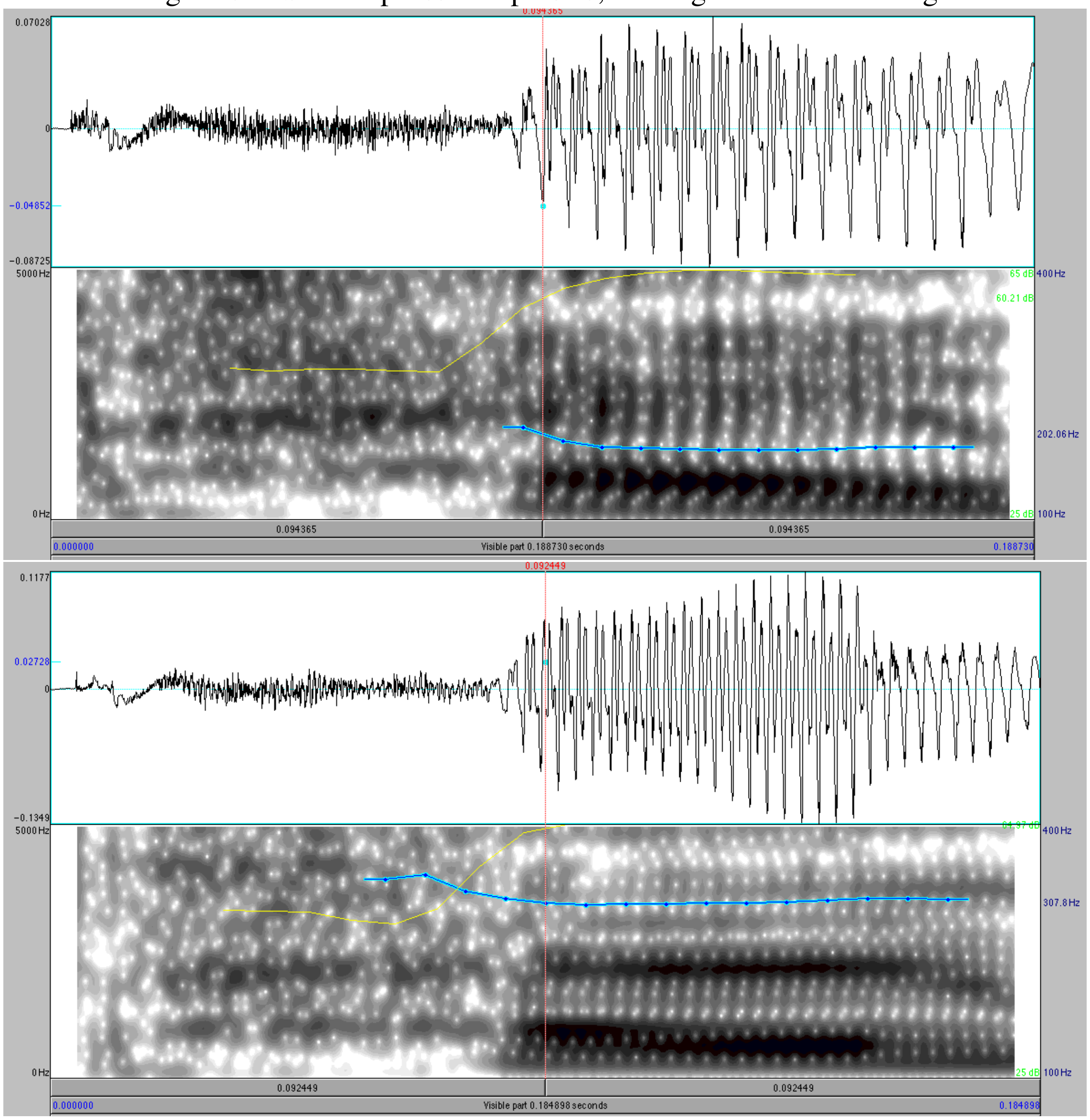
Ideally I would have created stimuli that were entire utterances, or at least entire word tokens. For example, the approach I used in a previous pilot experiment was to take whole utterances, excise and modify the word of interest, and splice it back into the original recording. However, this is particularly problematic for the current experiment. Ideally I would like to have taken words with LHLH and HHLH (or even the shorter versions of LH and HH) tonal contours and modifying them to reverse the initial tonal contrast. As demonstrated in other chapters however, the second position H is effectively upstepped in

H-APs, so modifying the pitch of the first syllable is not sufficient. An alternative approach would be to raise or lower the pitch of the first two syllables, or of the entire word. The problem with this approach is that in H-APs there is often a level or close to level contour over the first two syllables, whereas the relative rise in the contour for L-APs is drastic. There is no *straightforward* manipulation of pitch to transform between the two AP types. Therefore, I decided to use single syllable stimuli, taken from the initial syllables of APs.

The second problem I faced in devising the stimuli was the trade-off between natural variation and artificial variation. My stimuli were modified natural recordings, which by nature have variation which by hypothesis is ignored as non-contrastive phonetic variation by listeners. But since this variation is unknown and/or unpredictable, the results become more meaningful the more recordings that are used, because you can presume to abstract away from idiosyncracies that might have mislead, or informed, the listeners. On the other hand, the test characteristics, VOT and pitch, are being chosen by me through manipulation of the recordings, and more variation is also better here because the results become finer grained. In my pilot experiment I simply modified pitch from high to low or vice versa, but in this experiment I wanted intermediate levels. Since both natural and artificial variation are desireable, this presents a trade-off because another goal was to keep the data set relatively small due to real-world contraints. The key to a relatively large data collection in a relatively small amount of time was in keeping the recording and listening tasks to a small size.

Ultimately, I sacrificed natural variation: my stimuli were based on only two recordigns, one containing /p/ and one containing /p^h/. Specifically, I excised two syllables, /pal-/ and /p^hal-/ from instances of the words /pal-mok-til-il/ and /p^hal-mok-til-il/. I chose to use the recordings of speaker f08, who was one of the earlier participants and who I had established had no VOT contrast (in fact, a reversal of contrast). Speaker f08 also had double the number of tokens to choose from, and after scrutinizing those tokens I chose two that were as similar to each other as possible, and had extremely level pitch contours. In other words, the tokens differed drastically in terms of absolute pitch, but were nearly identical in terms of pitch contour, VOT, intensity contour, and general appearance of the spectrograms. I'll refer to these two original recordings as p-185 and ph-310: the number refers to the approximate pitch level.

Figure 6.1: Stimuli p-185 and ph-310, the original seed recordings



I then created a 6x6 stimuli space, varying both pitch and VOT over 6 discrete levels.

By using each original token as a seed for this space, 72 stimuli were created, the two originals plus 70 modified versions of those originals. The pitch ranged between 185 and 310 Hz based on the pitch of the originals, and the VOT ranged over the *relative values* of -45 to +30 ms, which was roughly an absolute range of 35 to 110 ms. Table 6.1 shows the stimuli space with the positions of the original recordings.

Table 6.1: Stimuli Space

	Relative VOT					
pitch	-45	-30	-15	+/-0	+15	+30
310				p^h		
285						
260						
235						
210						
185				p		

The stimulus ptich was modified using Praat's *change gender* function, with the formant ratio set to 1.0 so that formants would not be modified. This function utilizes the PSOLA pitch modification algorithm. When using this function, a target median pitch is given for the output, and that target pitch is what is used as an index into the stimuli space.

The two original recordings had very close VOTs, 80 ms for p-185 and 77 ms for ph-310. You can see that the pitch track indicates a much shorter VOT for ph-310, but my estimates were based primarily on the waveform. Since VOT is somewhat hard to define objectively, the stimulus VOT was based on the relative measure you see above. Reference points slightly before the onset of voicing were selected in each of the original recordings. For the negative relative VOT values, portions of the recordings were excised from the reference point backwards. This method preserves the stop bursts with room to spare. For positive relative VOT values, the 15 ms portion immediately prior to the reference point was copied and inserted at the reference point, twice for the +30 values. Pitch modification was performed first, and then those 12 +/-0 VOT stimuli served as input for the VOT modification procedure.

Each speaker was presented with a list of 74 stimuli: first, p-185; second, ph-310; third, all 72 stimuli, including p-185 and ph-310, presented in random order. The speakers were unware of the special nature of the first two stimuli. The rationale for this list structure was to provide a very minimal opportunity for listener acclimation to the voice of speaker f08. One speaker actually asked me if she could change her answer to the first stimulus, since after hearing the "contrast" of the second stimulus, she realized that the first was in fact a /p/. Each speaker was given a list of 74 sounds in Praat, and an answer sheet for which they had to circle either *pal* or *phal* (in Hangul) for each stimulus (forced choice). They were free to replay the sounds and move around in the list however they saw fit.

6.2 Results

Let's start with two extreme cases. First, speaker f20, born in 1955, considered every single token to be aspirated. I should note that my stimuli, not by conscious intent, are biased towards examining speakers who already show evidence of a change in aspiration contrast, because even the stimuli with the least aspiration have a VOT of around 35 ms. Nevertheless, f20's response was striking. On the other extreme we have speaker m06, who appears to give no regard to aspiration in his categorization, while relying heavily on pitch, as seen in the histograms in Figure 6.2.

Just to emphasize the result and illustrate my procedure, I created the same graphs after separating the data by original seed stimulus. In principle, either set of 36 stimuli would have been sufficient for the experiment, but using both sets of 36 I felt was important not only for balance, but because the pitch modification procedure does add an unnatural quality proportional to the degree of change. However, note that these graphs are mostly predictable from the combined graphs: for the four bars in the combined plot by pitch that are 100 percent one way or the other, we know that the original seed recording had no effect. This will be largely true for the other speakers as well, that there is little room for this sort of effect in the results, so I won't break down the results this way for other speakers. Here, for speaker m06, separating by seed recording, we can see that those intermediate tokens tended to be "misidentified" rather than identified as the same category as the original token. The histograms marked as h+ represent the tokens based on ph-310, and the ones marked as h- represent the tokens based on p-185.

Figures 6.2, 6.2, and 6.2 show the perception results by pitch for the other speakers. I felt it was more meaningful to show the results on an individual basis to see how robust the pattern is. High pitch appears to be a universal indicator for /ph/: all the 285 and 310 tokens were identified as /p^h/, with one exception, and 10 speakers also identified all the 260 tokens as /ph/. 7 speakers identified all 185 tokens as /p/, but in general the lower pitches resulted in more uncertainty than the higher pitches. To be clear about the exception mentioned above: one instance of one token at the 310 pitch level was identified as /p/ by speaker f21; see that speaker's histogram.

Figure 6.2: Perception results for speaker m06

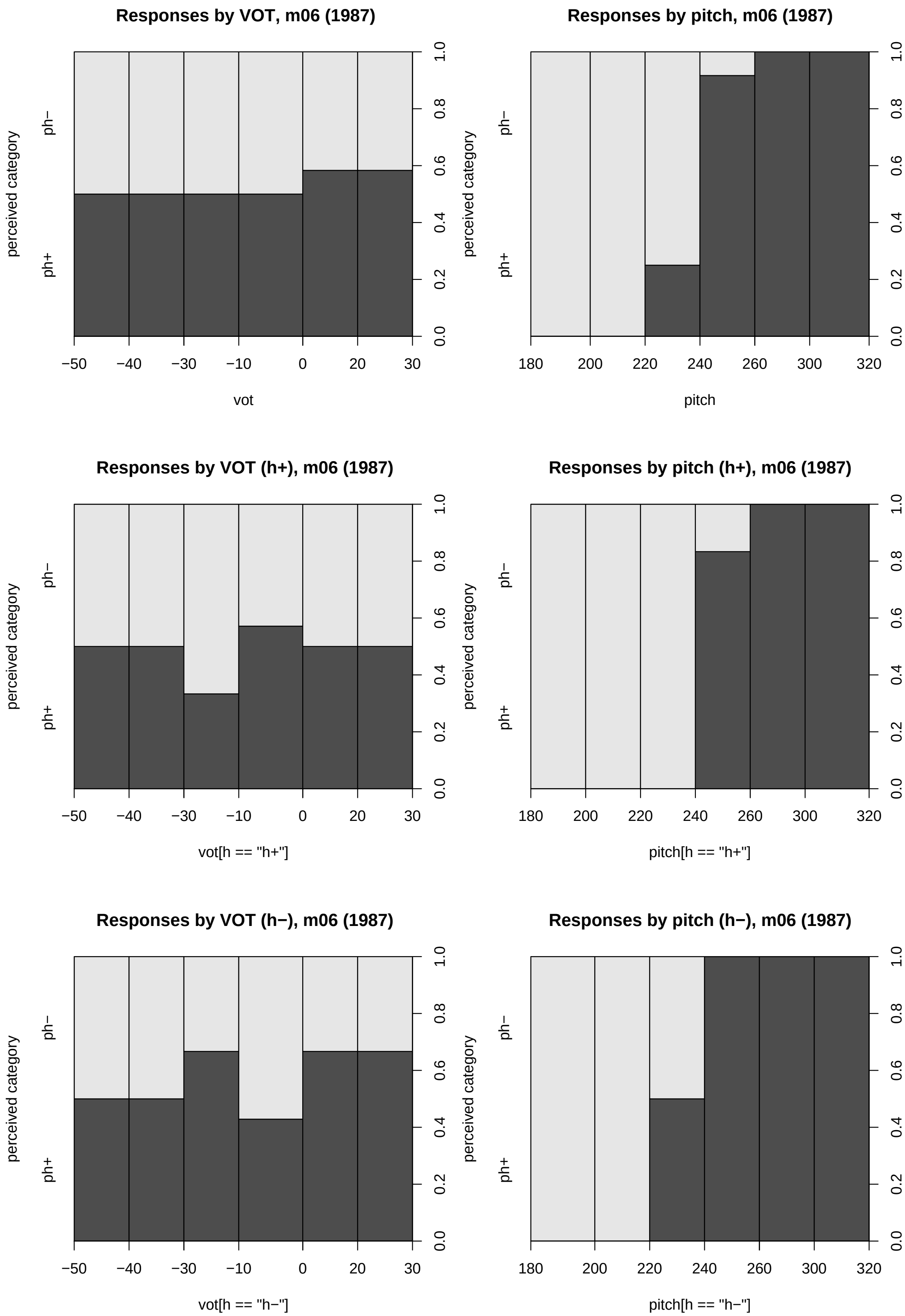


Figure 6.3: Speakers' perception by pitch, part one

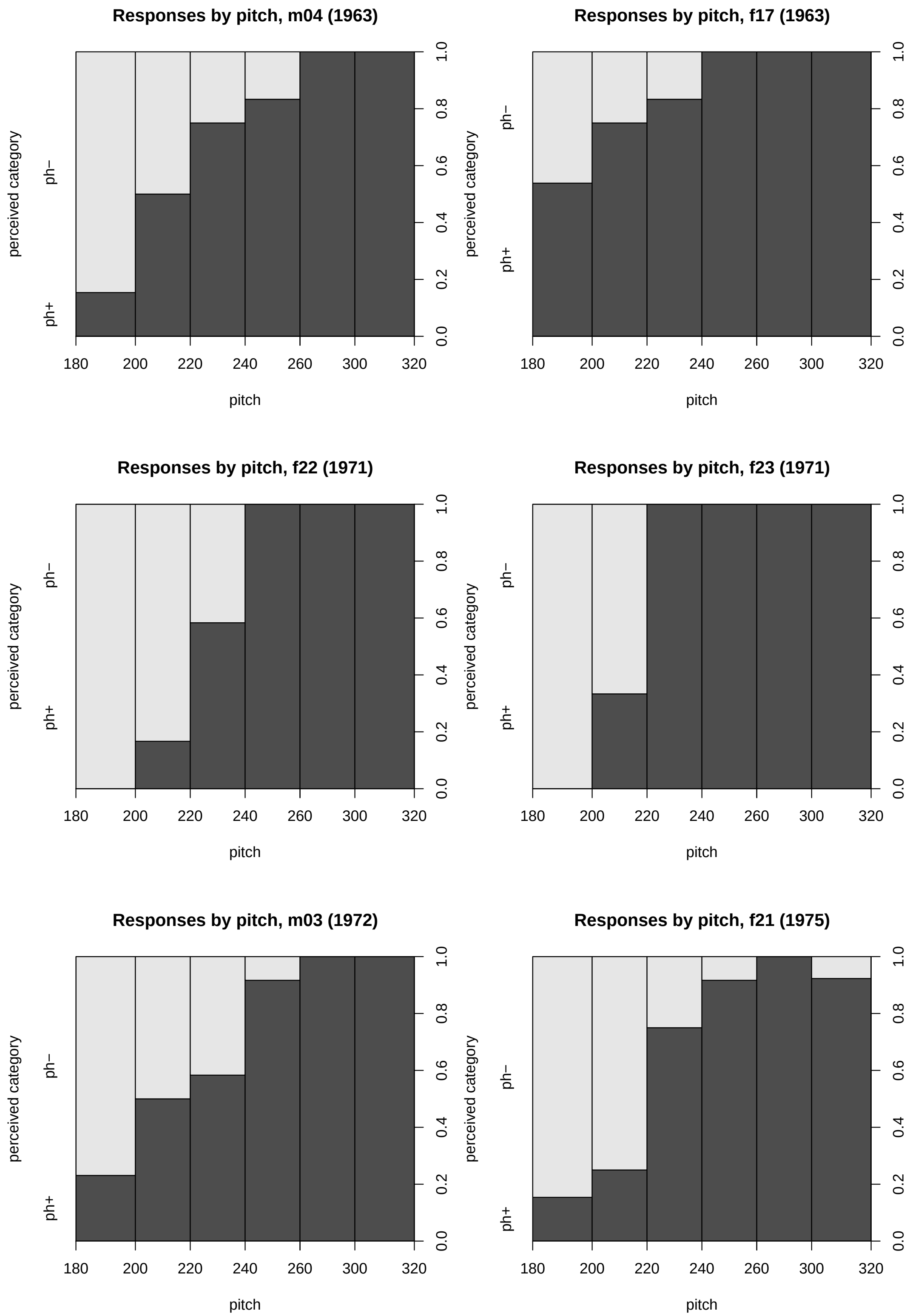


Figure 6.4: Speakers's perception by pitch, part two

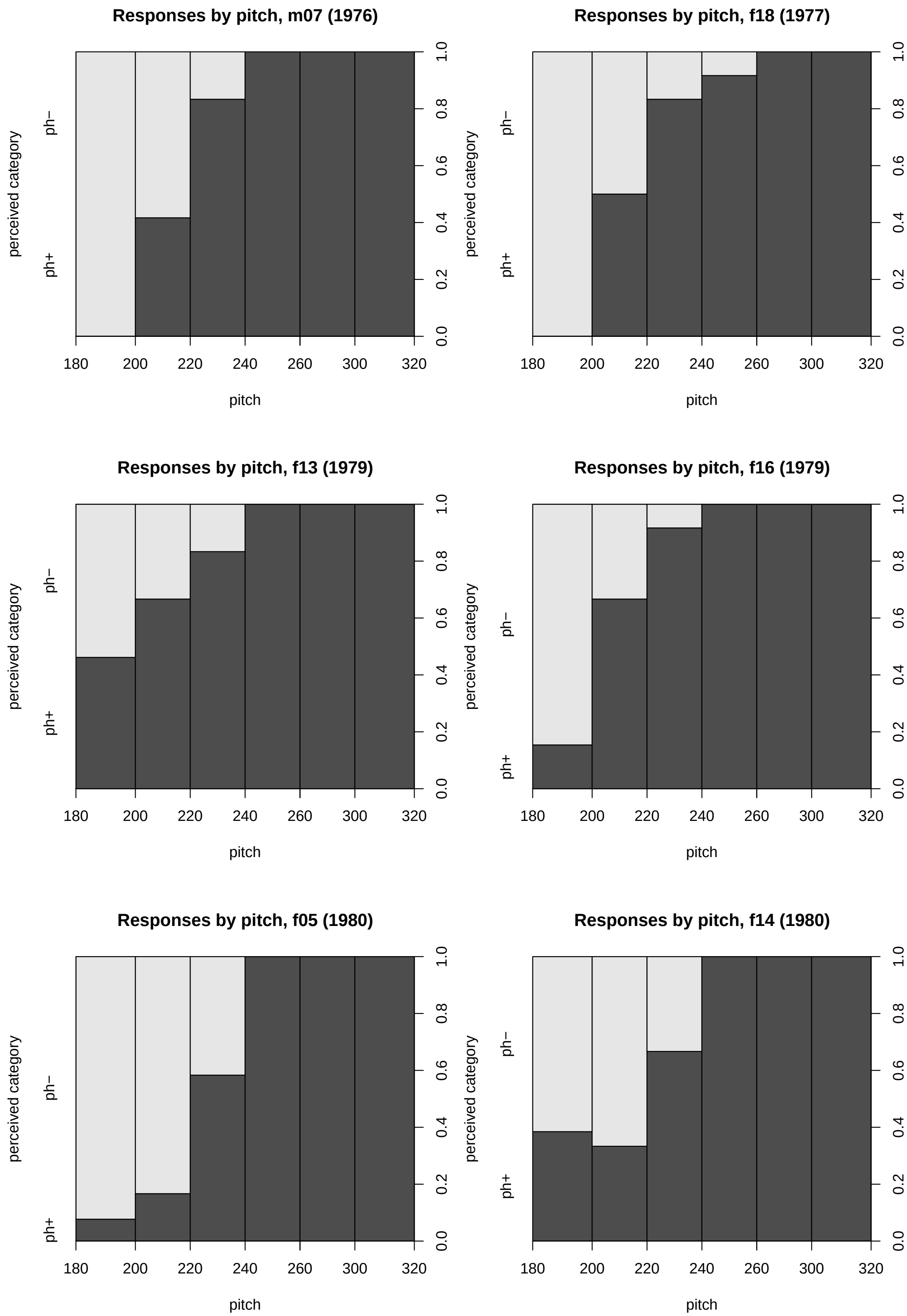
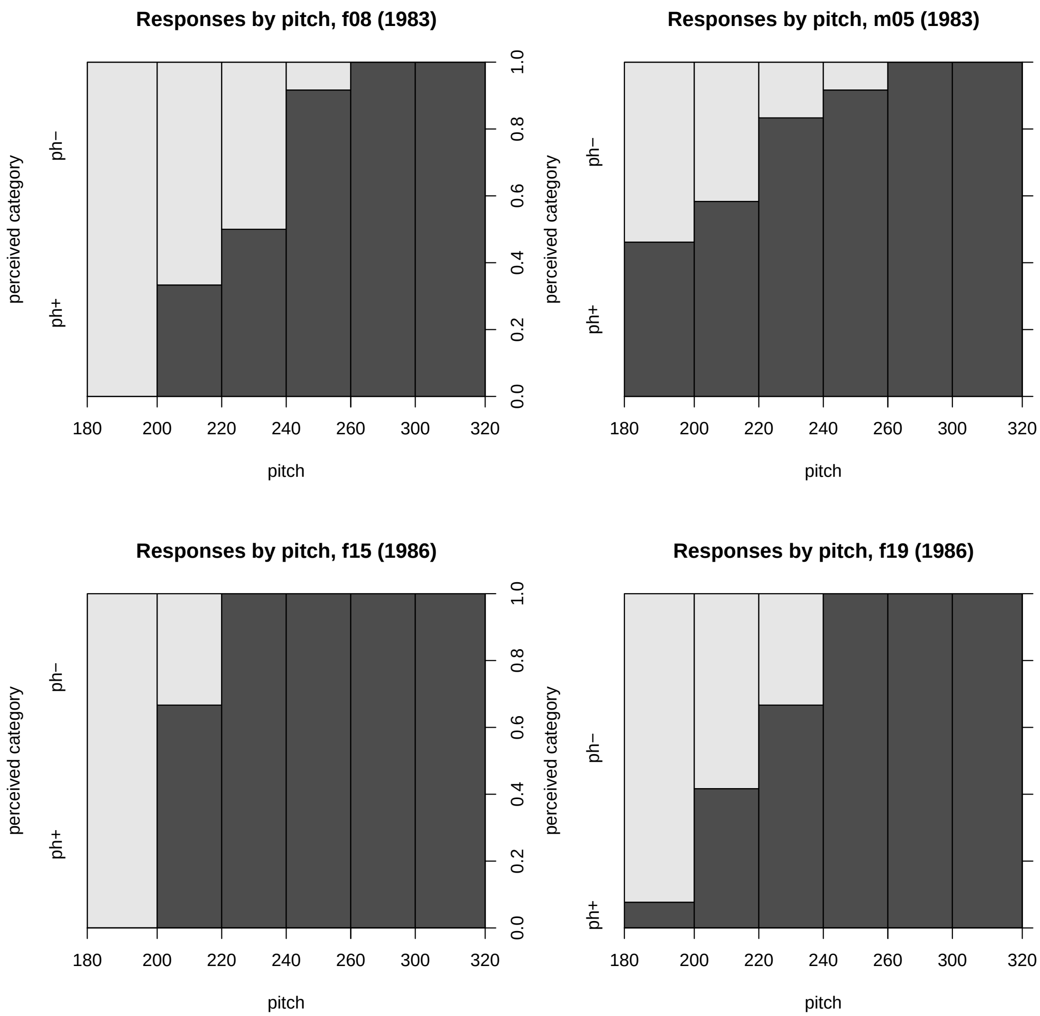


Figure 6.5: Speakers's perception by pitch, part three



Ideally the next thing to look at would have been a break down of those lower pitch levels according to VOT to see if, at intermediate or lower pitches, a perception of VOT took over. Unfortunately, after that breakdown, each bar in the histogram would only represent two tokens, which isn't particularly meaningful. I will address this in future work; see the conclusory chapter for discussion.

Although I'm hesitant to merge speakers when there's a change in progress, Table 6.2 gives a picture of the perception space by totaling all /p^h/ responses and all /p/ responses, and subracting the latter from the former; therefore negative numbers represent /p/ responses. From this table one can make out a rough line dividing the perception space. I also give the perception space for speaker m06 alone, in Table 6.3.

Table 6.2: Perception Space, combined responses

	Relative VOT					
pitch	-45	-30	-15	+/-0	+15	+30
310	28	28	28	–	28	28
285	28	28	28	28	28	28
260	20	26	24	28	28	28
235	-4	2	12	18	22	24
210	-26	-8	0	0	8	14
185	-28	-26	-22	–	-6	-14

Table 6.3: Perception Space for speaker m06

	Relative VOT					
pitch	-45	-30	-15	+/-0	+15	+30
310	2	2	2	3	2	2
285	2	2	2	2	2	2
260	2	2	0	2	2	2
235	-2	-2	0	-2	0	0
210	-2	-2	-2	-2	-2	-2
185	-2	-2	-2	-3	-2	-2

6.3 Conclusion

The perception experiment provides strong evidence that pitch dominates VOT in determining contrast being plain and aspirated segments. It's difficult to identify any sort of trend in this data, beyond the extremes of speakers f20 and m06. I would hope to see less uncertainty among younger speakers than older speakers for the lower pitch values, but this doesn't seem to be the case. The experiment may not be precise enough to bring this out. But it did identify the certainty among all speakers for higher pitch ranges: higher pitch means aspirated. We can also hypothesize that speaker m06 represents the vanguard of emerging perceptual distinctions: for him, pitch has completely taken over the functional load of the contrast.

Chapter 7

Tone and Pitch

Korean tone as it relates to this dissertation has two important elements: the contour of the AP and the tonal contrast between segment types, which I've been calling segment-induced tone. This chapter examines these properties through the measurement of pitch, then considers the history of tone in Korean and the implications of that history for the modern language.

7.1 Evidence from the Main Data Collection

Segment-induced tone was already present in the data from Han and Weitzman (1970), whose speakers were likely born in the 40's if not previous. I have not expected to find any sort of trend regarding tone since it's a well established phonological contrast. While there doesn't appear to be a trend in my data, it's also true that my speaker pool doesn't include enough older speakers to rule out a trend.

Using the recordings discussed in Chapter 4 I made pitch measurements for the words *pal-il*, *p*^{h}*al-il*, *pal-mok-til-il*, and *p*^{h}*al-mok-til-il*, taking a single measurement from the first syllable in each case. The four syllable words reflect the full AP patterns of [HL]HLH while the two syllable words are simply [HL]H. The shorter words turned out to be more consistent in terms of measuring the first syllable pitch, partly because the first syllable would be a little longer, and also because after resyllabification it had no coda. But for both pairs, 2-syllable and 4-syllable, I examined the pitch means and boxplots of the distributions. With the exception of the oldest speaker, f20, all speakers had quite distinct, non-overlapping distributions of pitch for the two stop manners. Speaker f20 showed a slight overlap of distributions, which gets back to the question of a potential difference between younger and older speakers, but again my sample doesn't sufficiently cover older speakers. A slight overlap in one speaker isn't particularly meaningful.

Table 7.1 shows the difference in mean pitch for each speaker using only the 2 syllable words.

Table 7.1: Differences of mean pitch for /phal-il/ and /pal-il/

speaker	YOB	pitch diff
f20	1955	36
f17	1963	54
f22	1971	54
f23	1971	93
f09	1972	81
f21	1975	123
f18	1977	161
f13	1979	93
f16	1979	78
f05	1980	60
f14	1980	103
f08	1983	107
f01	1986	62
f15	1986	78
f19	1986	78

speaker	YOB	pitch diff
m04	1963	42
m03	1972	46
m07	1976	40
m05	1983	41
m06	1987	30

Speaker f18 had an unusually large difference of 161 Hz. Her utterances were also particularly clear and consistent in terms of pitch contours, so I've chosen her data for illustration. Following are spectrograms and pitch tracks of 10 of her utterances containing the words /pal-il/ and /p^{h}al-il/, 5 each. The target words have been delimited. Remember that each of these target words forms a single AP, with a LH or HH pattern; the difference between the L-APs and the H-APs is unmistakable in the pictures. The preceding words, /na-nin/ and /ne/ each form an L-AP of their own. The pitch of the verb at the end of the utterance is normally degraded due to its boundary position, but is technically an L-AP, beginning with the segment /t/.

The H-APs for this speaker demonstrate, that at least for some speakers, there is a rise after the initial H, so that the HH contour is effectively upstepped. I have seen this effect often for both HH and HHLH APs, but the pattern is not consistent across speakers: sometimes H-APs have a level or even falling contour across the first two syllables.

Figure 7.1: L-APs from speaker f18

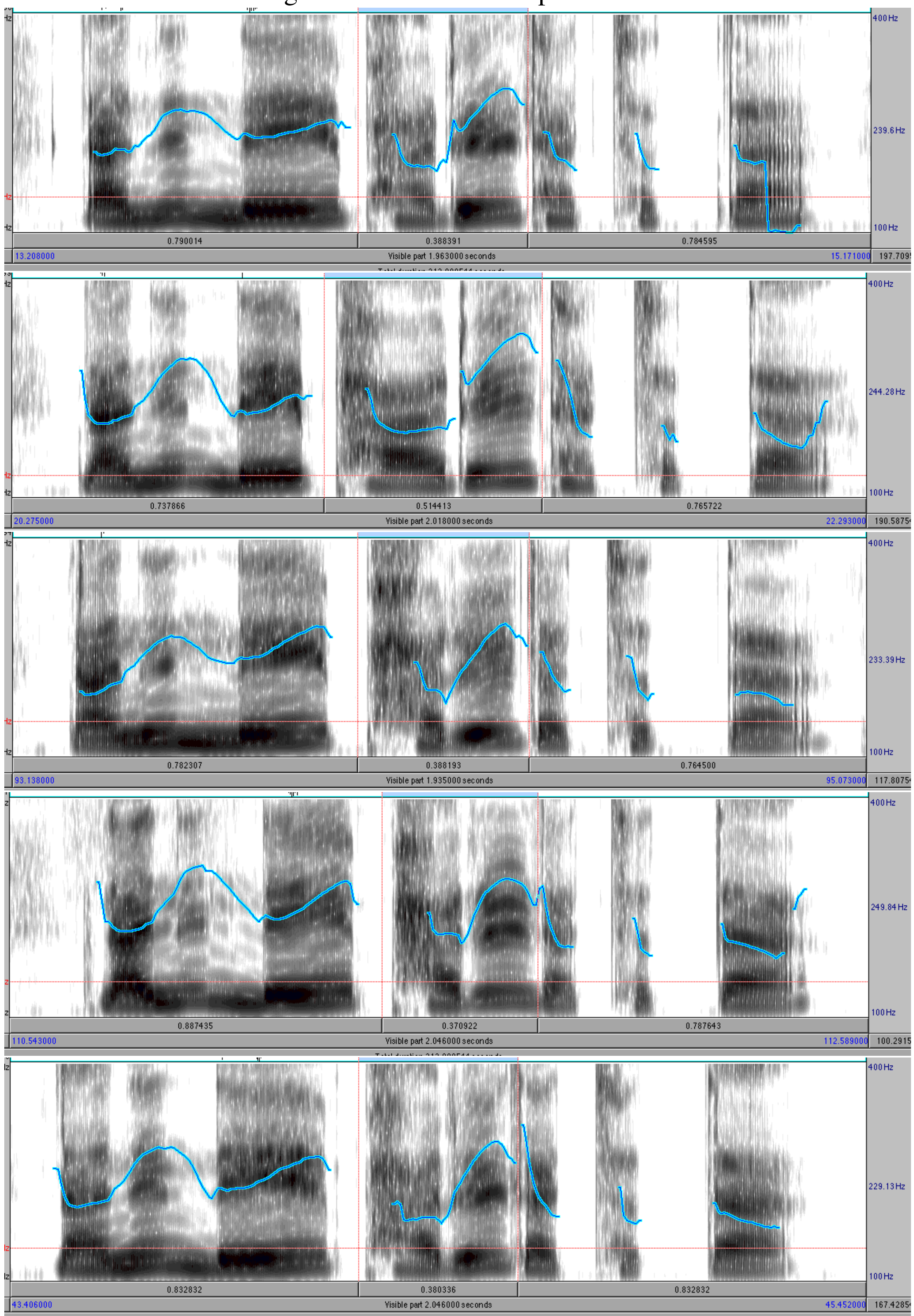
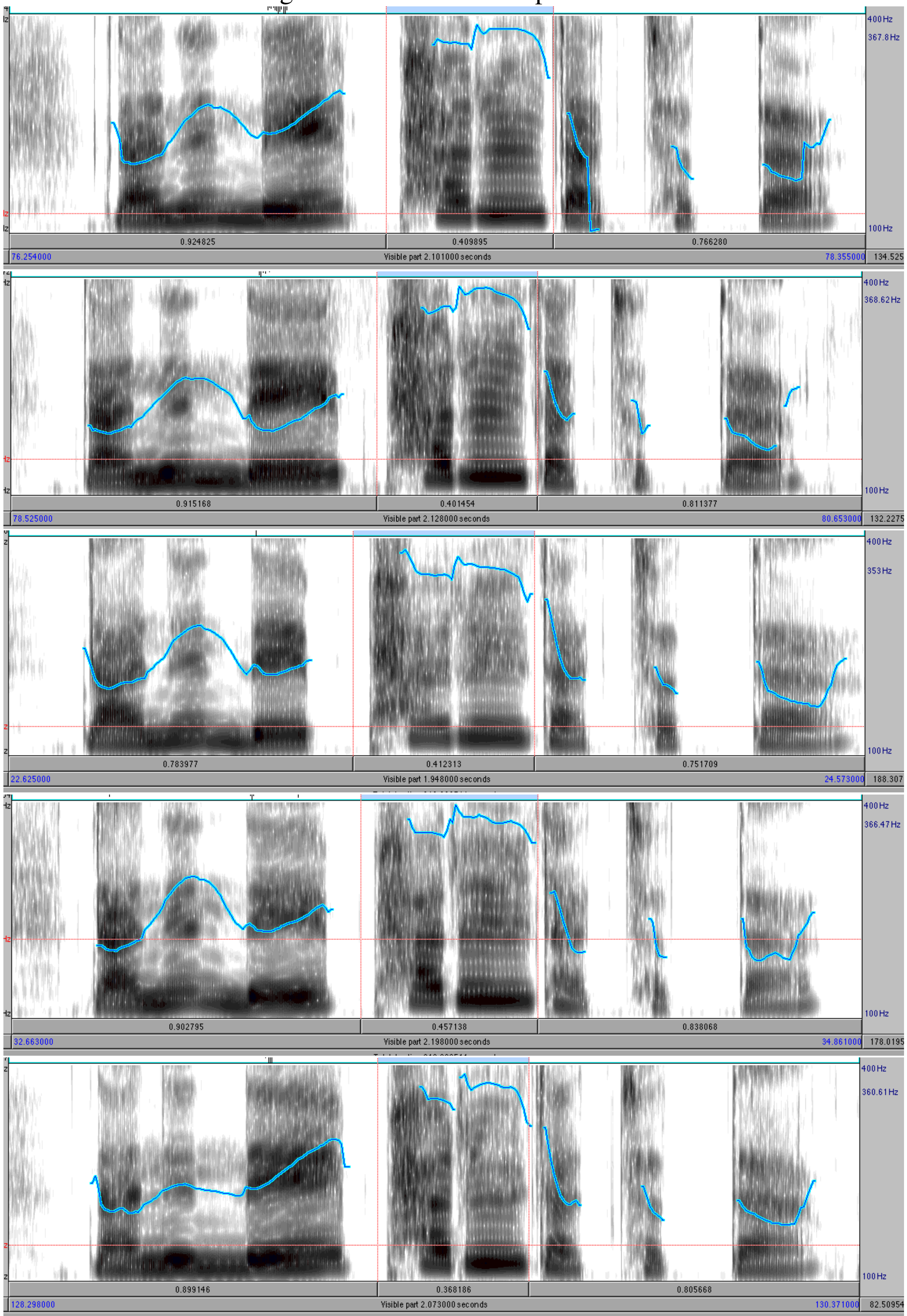


Figure 7.2: H-APs from speaker f18



The beginning of each utterance in this data set is *na-nin ne*, a two word sequence, each word forming a separate L-AP, therefore each with a pitch minimum at the beginning of the word. This is certainly true for the first word, and usually true for the second word, although less clear due to contextual forces. I measured these two minimums for speakers f18 and f08 and compared the values to those measured for *pal-il*.

Table 7.2:

	f18				f08			
	na-nin	ne	pal-il	pal-mok-til-il	na-nin	ne	pal-il	pal-mok-til-il
mean	196	229	195	195	190	184	181	193
SD	8.54	18.38	8.23	7.35	9.36	9.17	7.45	4.42
N	77	77	10	10	160	160	20	20

The means for *na-nin* and *pal-il* are not significantly different ($p > .1$) for speaker f18, but they are for speaker f08 ($p < .0001$). Since the two speakers showed different patterns in this respect, I included the measurements for *pal-mok-til-il* which I consider to be less reliable. For speaker f18, this mean is right in line with the others, while for speaker f08 it is significantly higher than that of *na-nin*.

So it's difficult to establish a relationship between the L that begins the utterance and the L that is essentially the focal tone of the utterance, that begins the target token. Such a relationship would potentially be dictated by two factors: the differing phrasal position, and the difference in segment type. The next section details these issues.

7.2 Nonsense Words, Trial 1

I collected a small amount of data based on nonsense words, first using the words pal-mok-til-il and p^hal-mok-til-il as models. Taking the sequence al-mok-til-il as a base, I added 9 different consonants to the beginning: /p, p^h, p*, t, t^h, t*, n, m, y/. Including one case where no onset was added, this yielded 10 different tokens, and a list of 30 tokens was formed by repeating each token 3 times and randomizing the order. No carrier sentence was used with these tokens. Speakers f08 and m03 were recorded reading this list. Interestingly, with a single exceptional measurement, each speaker's distribution could be divided into non-overlapping distributions by separating the plain stops from the sonorants.

Table 7.3: Mean pitch for L-segments, plain stops vs. sonorants

	f08			m03		
	all	plain stops	sonorants	all	plain stops	sonorants
mean	180.4	185.8	177.8	112.2	116.7	110.0
SD	5.12	3.49	3.36	3.98	1.75	2.63
N	18	6	12	18	6	12

The narrow difference in means suggests a microprosodic effect, in the sense of automatic phonetics, and conforms to the standard microprosodic pattern where voiceless segments result in higher pitches[1]. Among the H-segments there is no clear measuring point as with the L-segments where there is almost always a local minimum. Sometimes the contour following an H-segment is quite level, sometimes slightly rising or falling, etc. For these tokens I took an average pitch value over most of the first syllable, and then averaged those values. Separating the values into tense and aspirated distributions did not yield a significant difference, therefore for all H-segments, speaker f08 had a mean of 246.2 and a standard deviation of 18.1, and speaker m03 had a mean of 142.3 and a standard deviation of 6.45 (N=12). My data set isn't large enough to rule out microprosodic effects here too, but it's likely that if there was a phonological difference, i.e. two different H tones, that difference would be apparent even here.

The lower value of plain stops here doesn't mesh with the results from the previous section, where plain stops occupy the target position of the utterances and /n/ begins the

[1]Some readers may object to me resorting to microprosody here after going on about how it can't explain segment-induced tone. After all, this effect here is arguably non-local also, with the pitch minimum sometimes occuring in the middle of the syllable. The case against microprosody being involved in segment-induced tone is really based on both non-locality *and* high degree of difference. Segment-induced tone is clearly intentional, in the phonological sense.

other positions. When compared to the initial position value, the target words are the same for speaker f18 and lower for f08, rather than higher as with the nonsense words. An alternative interpretation is that prosodic position is an overriding factor, that the L in the target word is lower relative to the initial L of the utterance due to its position, at least for speaker f08.

7.3 Nonsense Words, Trial 2

Taking a different approach, I created these 7 nonsense words: pap-il, p^hap-il, p*ap-il, hap-il, map-il, wap-il, ap-il. These words were inserted into 3 different carrier phrases that differed only in the pre-nominal modifiers used: ne, se, and ne se together, which means "my new". Note that /s/ is an H-segment. The carrier phrase, with both modifiers, was *na-nin ne se — ta-cho-ta*, "I like my new —."

7 words times 3 carrier phrases times 2 repetitions yielded 42 sentences, randomized, and read by speaker f08. Pitch values were measured for each syllable up to and including the first syllable of the verb, after which the low boundary tone affected the contour. The exact method of measurement varied by syllable however, based in part on whether the syllable was AP-initial or AP-final. The pronoun na-nin and the target word in object position each formed two syllable APs, so had one syllable of each type. The first syllable of the verb was an AP-initial syllable. Both pre-nominal modifiers formed their own APs, so those syllables were both AP-initial and AP-final. AP-initial syllables with L-segments usually have a local minimum, and in those cases that minimum was measured. AP-final syllables tend to rise to a final peak, and those cases that peak was measured. The pre-nominal modifiers sometimes had both local extrema, and those cases the initial minimum was measured, however they were often just level pitch. For syllables that had level or falling pitch, which was often the case with syllables with H-segments, the pitch value was

taken from the middle of the vowel.

The first pattern to note is that of the pitch of the first syllable, which was always /na/: the mean pitch is 184. This exactly matches the value found for /na/ in the previous full sentence data for speaker f08. Now let's consider the pitch of the first syllable of the target word, where the first segment was varied. The mean pitch for L-segments only was 161. This mean is significantly lower than that of the first syllable, both statistically ($p << .0001$) and qualitatively.

This data set was designed in part to see how pre-nominal modifiers, or more specifically the tonal values of pre-nominal modifiers, might affect the target syllable in different ways. The word *ne*, forming an L-AP, has a L or LH pattern, while the word *se*, forming an H-AP, has a H pattern, normally higher than the H that ends an L-AP. Splitting the target syllable values into three groups based on modifier type yields three distributions that are not significantly different ($p = .4, .5, .8$), therefore this L-lowering seems robust. For this calculation N = 10, 5 L-segments times 2 repetitions. Therefore the following breakdown is justified given that the data set is small but that we don't need to control for pre-nominal modifiers.

The relative pattern here is the same as for the previous full sentence data, but with a higher degree, that is, 161 is more of a decrease for the target position L. This again has two possible interpretations, a prosodic position effect, or an effect due to the fact that this nonsense word data includes sonorants which are potentially lower than plain stops. Note that I am not seriously considering a third possibility that nonsense words have an effect. The 6 plain stop values average to 174, and the 18 sonorant values average to 157. Therefore it seems that there is *both* a segment effect and a prosodic effect.

Now consider the H-syllables. In this data set since the repition was only 2, the data can't be broken down in every dimension at once. When dividing H-segment pitches by manner/segment into three groups, /p^h, p$*$, h/, there was no significant difference between the means ($p = .14, .24, .97$). When grouping all H-segments together and dividing by prenominal type, there was also no significant differences, although one category was almost significantly higher than the other two ($p = .08, .06$), so more data might have shown a significant difference. The pitch mean for the target word was 278 after *se*, but only *se*. Following *ne* the mean was 255, and following both, *ne se*, the mean was 258. The effect I was looking for here, for which there is slight evidence, is that an H could cause a following H to upstep. But again, the numbers in this data do not fall below the $p = .05$ significance level.

Finally, the pitch for the two syllables in the target word when the target word forms an H-AP did not show a significant difference in means (263, 267, $p > .1$). This suggests that H-APs are not upstepped versions of L-APs, but that the initial and final Hs are equivalent. In fact, these means belie the fact that often there is a pitch fall in these H-APs.

7.4 Summary

The nonsense word trials looked at a broader range of possibilities than the change in progress data, but unfortunately I did not commit to a high sample size, which I feel weakens my results. Particularly when it comes to the possibility of upstepping, which sometimes seems to occur in the spontaneous corpus data. However, I believe my results here support that, at least for speaker f08, there is both a downstepping effect, of some sort, and a microprosodic difference among L-segments based on the obstruent/sonorant distinction. That is, all L-segments belong to the same class in terms of segment-induced tone, despite there being microprosodic effects. The results are less clear for the H-segments, although it seems likely that they also form a single class phonologically. Other authors have reported lower pitch values for the tense series relative to the aspirated series, which I suspect is a microprosodic difference. The sample size in my trials may have been too low to bring out such an effect.

Chapter 8

Tonogenesis

There are two tonal phenomena in Korean: the assignment of phrasal tones, and segment-induced tone. The change in progress in Seoul Korean suggests that segment-induced tone could be an early stage of tonogenesis. In this chapter I walk through several possibilities for tonogensis in Korean, before concluding that modern Seoul Korean has had lexical tone since the time of Middle Korean.

8.1 Defining Tonogenesis

Briefly we should ask, what is tonogenesis really? Obviously the origin of lexical tonal contrasts should count, but what about the origin of Japanese pitch accent, or the origin of English stress accent? While *stressogenesis* or *metricogenesis* should be equally interesting and broadly similar to tonogenesis, I'm excluding English-like stress systems from the present discussion and from the domain of tonogenesis, mainly for simplicity's sake. While stress often (or always?) involves pitch phonetically, it can be considered a distinctly different type of system than a lexical tone system. Stress is a system of relative prominences, for which some languages, like English, have a lexical contrast. In addition, as Hyman

(2006) defines it, the lexical accents in a stress system are obligatory and culminative: each word has exactly one maximal stress. Finally, stress accents aren't inherently high tones, and low tones may anchor to stress accents in the right circumstances.

However, I will consider tonogenesis to cover all other tonal systems, including for example the so-called pitch accent of Japanese. While the analysis of Japanese doesn't require underlying tones, it is nevertheless a system of lexical, tonal contrast. In other words, the system of contrast is both tonal and lexical, even if a particular analysis doesn't posit lexical tones. Tonal systems broadly construed, unlike stress systems, should not dispaly *both* obligatoriness *and* culminativity, while perhaps having one or the other. At least for the current discussion of tonogenesis, we can assume the previous generalization is correct.

In sum, we can define tonogenesis as *the evolution of a system of contrast that is both lexical and tonal*.

8.2 The Synchronic Status of Korean Tone

Descriptively, Jun's theory of the AP seems correct and well-established, but we can still question its phonological status. In Jun's theory, the assignment of tones is based on syllables or moras. This is the major contrast between the Korean AP and the Japanese AP, the latter of which crucially relies on the presence of lexical accents. It's at least possible that Korean is actually the same as Japanese in this respect, except that the accent is manditorally in second position in Korean.

Segment-induced tone raises more important questions. Jun's description of this phenomena is that the type of the AP-initial segment dictates which of the two possible contours is assigned. Another way of analyzing it would be to say that the L-AP contour is always assigned, and that there's a rule where H-segments change the initial L to an H. In other words, a tonal alternation is conditioned by melodic segments. So far I have found no other case where such a rule has been proposed. However, there are at least 3 cases where the reverse has been proposed, where tone conditions a segmental alternation (Bradshaw, 1999). All three cases, from three unrelated languages, involve an L tone triggering voicing in obstruents. In addition, *depressor consonants* are widely attested in African languages, where a class of segments effects the placement of H tones (Bradshaw, 1999). Korean segment-induced tone could perhaps be given a similar analysis, but the simple description of it that I gave above doesn't seem to fit the pattern of other attested consonant-tone interactions.

On the other hand, in cases of attested tonogenesis, there may be transitory stages with synchronic phonologies similar to Korean, with segment-induced tone. The issue of the synchronic representation for Korean segment-induced tone can't be completely separated from theories of tonogenesis, which I turn to below.

8.3 Tonogenesis Pathways

Consider the most often mentioned tonogenesis pathway: the reinterpretation of microprosodic differences between initial obstruents leads to an L/H contrast correlating to the voice/voiceless contrast, the latter being subsequently neutralized to voiceless. Assuming the reanalysis and neutralization take place as separate steps, there would be an intermediate stage with redundant contrasts, giving a pathway as follows:

1. Voicing contrast only.

2. Reanalysis of microprosody results in redundant tonal contrast.

3. Voicing contrast is neutralized (to voiceless).

This description leaves open the question of the synchronic phonology of the intermediate stage, which is similar to the question of Korean's synchronic phonology. Consider the pathway below:

1. Voicing contrast only.

2. Voicing conditions redundant tonal contrast on the surface.

3. Voicing and tones exist redundantly in the underlying representation.

4. Voicing contrast is neutralized (to voiceless).

This four stage pathway separates the phonologization of microprosody from the lexicalization of tone. However, a particular theory of phonology may rule out one of the intermediate stages. If your theory didn't allow segment-induced tone, explicit in stage 2, you could eliminate that stage such that phonologization and lexicalization occur simultaneously. On the other hand, stage 3 contains a lexical redundancy, which is theoretically distasteful; stage 3 could be eliminated by allowing lexicalization and neutralization to occur simultaneously. There are further possibilities for this pathway if your theory of phonology allows the tone-induced voicing described in the previous section. Consider the following fully articulated pathway:

1. Voicing contrast only.

2. Segment-induced tone.

3. Lexical redundancy.

4. Tone-induced voicing.

5. Tonal contrast only.

Any number of the intermediate stages could be eliminated, giving 7 possible sequences. Even if your theory of phonology allows all 3 of the intermediate stages, this doesn't guarantee that all 3 are present in the pathway. It's even possible that different individuals follow different sequences; all 3 intermediate stages are potentially identical on the surface.

The above 5 step pathway is in terms of synchronic stages, and can be represented as a 4 step pathway of diachronic changes, as below:

1. Phonologization of microprosody.
2. Lexicalization of tone.
3. Underlying neutralization of voicing.
4. Surface (complete) neutralization of voicing.

These processes must all happen, but not necessarily separately. If steps one and two occurred together, there would be no intermediate stage of segment-induced tone, tone would immediately become lexical. However, it seems unlikely to me that lexicalization could occur regularly in this situation. It seems more likely that the reinterpretation of microprosody as lexical tone would be extremely idiosyncratic: not only might different words receive different interpretations, different speakers might interpret the same word differently. Such a diachronic process wouldn't yield a uniform consonant-tone correlation. For a completely regular correlation between segments and tone, whether synchronically or diachronically, I think the phonologization of microprosody must occur separate from and before the lexicalization of tone.

Returning to Korean, if we just consider the synchronic evidence for Korean, we can't really tell whether or not tone has been lexicalized. Following the argument of the preceding paragraph, the regularity of the consonant-tone correlation in Korean would imply that segment-induced tone must have been present at some point, if not now. The question would then become, what would motivate lexicalization. Clearly the phonemic merger of plain and aspirated stops would motivate it, but putting that aside, could a stage of underlying redundancy have arisen from a stage of segment-induced tone without a phonemic merger? This would require learners to ignore a generalization that clearly they are capable of making, since in this scenario we are starting from a stage of segment-induced tone. Therefore I would argue no, that Korean has not lexicalized tone.

That is, I *would* argue that, if I thought Korean was participating in a microprosodic tonogenesis pathway as described above.

8.4 The History of Korean Tone

8.4.1 Overview

A cursory look at the literature on Korean would yield the following picture:

1. Middle Korean, an attested language, had lexical tone.
2. Standard Korean no longer has tone.
3. Some dialects of Korean, like Northern Kyungsang Korean, have retained lexical tone.

Statements like number 2 are interesting for a couple of reasons. First, few people have recognized the phrasal tonal patterns Jun has identified (Jun, 1996b, 1998). Second, the supposed lack of tone in Seoul Korean, along with the sound change currently underway, suggest the possibility of tonogenesis.

Some authors have described the Middle Korean tonal system as a three way tonal contrast, largely because that's the nature of the orthography. Middle Korean orthography, like the modern system, divided segments into syllabic units, but in addition there were marks for syllables with high tone or rising tone. Syllables with low tone were unmarked. However, the description of the system in Ramsey (1991) is that of an accentual system very similar to Modern Japanese: Middle Korean words can be analyzed as having an accented mora, or none at all, just as in Japanese. In Middle Korean, the accented mora carried the first high tone, after which the tones in the word are variable and not contrastive. Words with no accent had all low tones. Syllabes marked as having rising tone in the orthography can be analyzed as having a low mora followed by a high mora.

The goal of Ramsey (1991) is to propose a theory for the origin of the Middle Korean system: he believes Proto-Korean had a non-contrastive system of word-final prominence. However, this is beyond the scope of the present discussion, and I won't describe that theory for simplicity's sake. My goal here is to use evidence from that work to suggest that the tonal properties of Modern Seoul Korean originate in the system of Middle Korean. I believe it's an over-simplification to say that dialects like Seoul simply lost the accentual system.

8.4.2 Cross-dialectal Evidence

In hindsight, Jun's work should have been the first suggestion that there was something interesting going on historically, since Seoul and Chonnam have the same tonal phenomena. It's perhaps not particularly striking that they both have a phrase-initial LH pattern, but it is striking that they both have the same pattern of segment-induced tone, since this is a specific and somewhat odd feature. More importantly, Kim (1994) demonstrates that the same segment-induced tone pattern is present in Pusan, part of the South Kyungsang dialect, and Kenstowicz and Park (2006) demonstrate this for both North and South Kyungsang Korean. This is important because the Kyungsang dialects, along with South Hamgyong Korean, are recognized as having preserved lexical tone/accent. I'll return to these latter dialects below, but the point I want to make here is that segment-induced tone is a cross-dialectal property of Korean. Since the phonetic literature primarily uses Seoul Korean, it can lead to the impression that segment-induced tone is an innovation of that dialect, when in fact it likely goes back to a common ancestor of the dialects. The odd pattern of segment-induced tone makes parallel evolution or spreading through contact unlikely.

8.4.3 Relevant Facts about Middle Korean

Ramsey (1991) divides Middle Korean into eight classes. Of particular interest is Class 2b, all of which begin with either a consonant cluster or with an aspirated stop. Ramsey proposes that the proto-forms of these verbs can be reconstructed as follows:

*pìsítá	>	psítá	'use'
*pìHítá	>	p^hítá	'spread'

The process of syncope illustrated here was not unique to this verb class, and is part of Ramsey's theory of Proto-Korean tonogenesis. Accordingly, it is not specifically relevant to my theory of the subsequent history of Korean; this syncope occurred before Middle Korean. But what is striking about this verb class is that the verbs all begin with clusters or aspirated stops. Since the clusters evolved into the modern tense stops, this verb class began solely with the progenitors of H-segments. Furthermore, as can been seen in the above examples, the syncope process resulted in an initial high tone, which is also true for all verbs in Class 2b.

At this point, the numbers given by Ramsey become important:

1. Out of 472 verbs tallied, 50 are in Class 2b, and 46 verbs belong to Class 2a. These 96 verbs are the only verbs that uniformly begin with a high tone.

2. 28 verbs total, Classes 3 and 4, sometimes begin with high tones and sometimes low

tones, depending on their inflections.

3. The remaining 348 verbs always begin with a low tone.

4. All 50 verbs from Class 2b and at least 5 from 2a begin with clusters or aspirated segments. Although exhaustive word lists are not given, there are very few examples from other classes with these onsets.

Based on these numbers, there is a very high correlation among the verbs between initial H-segment progenitors and initial high tone. Modern segment-induced tone may be a generalization of this correlation. But before I continue this argument, we need to return to the dialects of Kyungsang and Hamgyong, that clearly preserve a tonal contrast.

8.4.4 Dialects that Perserve the Lexical Contrast

Ramsey (1975) discusses dialects that have perserved lexical tonal contrast from Middle Korean: Kyungsang, both North and South, and South Hamgyong. South Hamgyong can be analyzed just like Modern Japanese, at least in terms of the basic pattern: the first mora is low, unless it is accented; the accented mora is high; moras after the accented mora are low; any intervening moras between the first mora and the accented mora are high. Therefore a four mora noun stem has five possible realizations, shown below. The no accent class and the final accent class can be distinguised by the tone of a suffixed mora.

accent position	tonal realization	tone of a suffixed mora
$\mu\mu\mu\mu$	LHHH	H
$\acute{\mu}\mu\mu\mu$	HLLL	L
$\mu\acute{\mu}\mu\mu$	LHLL	L
$\mu\mu\acute{\mu}\mu$	LHHL	L
$\mu\mu\mu\acute{\mu}$	LHHH	L

When creating correspondences between South Hamgyong and Middle Korean, the modern accent positions correspond to the first high tone in Middle Korean. Therefore while the tonal realizations are different, South Hamgyong has perserved the accents of Middle Korean (with some exceptions of course). When adding the Kyungsang dialects to the correspondences, we see that these dialects show a leftward accent shift. In this analysis, morphemes, both stems and suffixes, can be pre-accented. This analysis is justified by the behavior of certain suffixes which can place an accent on the final mora of a stem. However, all that concerns us here is the possible tonal classes, illustrated below.

South Hamgyong and Middle Korean accent	South Hamgyong tone	Kyungsang accent	Kyungsang tone
$\mu\mu\mu\mu$	LHHH	$\mu\mu\mu\mu$	LHHH
$\acute{\mu}\mu\mu\mu$	HLLL	$\acute{}\mu\mu\mu\mu$	HHLL
$\mu\acute{\mu}\mu\mu$	LHLL	$\acute{\mu}\mu\mu\mu$	HLLL
$\mu\mu\acute{\mu}\mu$	LHHL	$\mu\acute{\mu}\mu\mu$	LHLL
$\mu\mu\mu\acute{\mu}$	LHHH	$\mu\mu\acute{\mu}\mu$	LHHL

The pre-accented class, when phrase-initial and therefore lacking a preceding mora for the accent, realizes the unusual pattern of high tones on the first two moras, followed by all low tones.

Finally, recall that Kim (1994) and Kenstowicz and Park (2006) have shown that these dialects have segment-induced tone, just as Seoul and Chonnam do. The initial H or L due to the accent class can each be divided into their own high and low region in tone-space, based on the presence of initial H- or L-segments.

8.4.5 Connecting Segment-induced Tone to Middle Korean

It would seem that segment-induced tone existed before the dialects diverged from each other. If South Hamgyong also has segment-induced tone, we could project the pattern back to Middle Korean. If South Hamgyong does not have this pattern, this might indicate an intermediate ancestor that excludes that dialect. This distinction will become relevant below. For now, let's just assume that segment-induced tone co-existed with accentual tone in Middle Korean, just as it does in Modern Kyungsang. Since segment-induced tone corresponds regularly with initial clusters and aspirates, it therefore correlates strongly with initial accent, because those segments correlate with initial accent, as described above. Therefore the evidence suggests that Class 2b verbs had high tone both from lexical accent and from segment-induced tone, independently.

At this point, if I just say that lexical accent was lost in Seoul, I haven't added anything to the discussion, other than pointing out that segment-induced tone can be projected back to an earlier stage of Korean. But consider the following:

1. In Seoul and Chonnam, segment-induced tone refers to a regular correlation between H-segments and an initial tonal pattern of HH, as opposed to the default LH pattern.
2. In Kyungsang, the initial HH pattern corresponds to initial accent in Middle Korean.
3. The evidence suggests that in Middle Korean, initial accent and the high tone of segment-induced tone were strongly correlated, due to a 100% correlation in a particular verb class.

I propose that the Seoul and Chonnam dialects conflated the two types of high tone, accentual and segment-induced, and in the process spread the HH tonal pattern to all words that began with H-segments[1]. In turn, words that began with L-segments but had initial ac-

[1] It should be obvious by now that I haven't accounted for the fact that /h/ is an H-segment. In fact, I don't know whether /h/ is an H-segment in Kyungsang, since Kenstowicz and Park (2006) only investigate stops.

cent, which appear to have been few in number, would have joined all other accent classes, which collapsed to a single class with a LH initial pattern. 288 of the 472 verbs and 160 out of 236 bimoraic nouns from Ramsey (1991) have a LH pattern, so it makes sense that this pattern would become the elsewhere case.

As a side point, my theory suggests that the accent shift was an innovation of an intermediate ancestor of Korean that excludes South Hamgyong, due to the initial HH pattern which is not present in South Hamgyoung. If it turns out that South Hamgyong doesn't have segment-induced tone, then segment-induced tone might be an innovation of the same ancestor, and might be somehow connected to the accent shift.

8.4.6 Implications for the Synchronic Analysis

History of course doesn't dictate a synchronic analysis, and theoretical assumptions could dictate that the regularity of segment-induced tone be captured derivationally. But consider the advantages of a lexical account. First consider Kyungsang Korean, and how that would be analyzed. While we could modify underlying tones in the derivation to produce the four way surface contrast that exists for initial syllables, it would be simpler to just associate the correct tonal contours to the underlying form to begin with. For Seoul and Chonnam, rather than a phrasal application of an LH contour, followed by a derivational change to the first tone conditioned by H-segments, the correct tonal contour (or accent) could be part of the underlying representation. This analysis would make the system of these dialects fundamentally the same as the other modern dialects, the same as Middle Korean, and the same as Japanese and other similar languages. Rather than consonant-tone interactions of a sort not attested cross-linguistically, we would simply need contraints on the underlying inventory to assure the regularity of the pattern. Maintaining this regularity somehow in the grammar is important, since loanwords and nonsense words obey it.

8.5 Conclusion

The historical and cross-dialectal evidence suggests that the two tonal phenomena of the Seoul and Chonnam dialects, the second position high and segment-induced tone, are both descendants of the Middle Korean tonal/accentual system. In short, the modern dialects, I argue, reduced multiple lexical classes to just two, characterized by initial LH and HH tonal contours. Even if a derivational analysis of the modern synchronic system is more desirable, my theory sheds some light on the origin of the system. It's still unclear how segment-induced tone arose, but it would seem very coincidental if the syncope process that resulted in so many word-initial accents and word-initial clusters didn't also have a role in its origin. Further investigation into the historical facts is necessary.

Furthermore, my theory suggests a reason for the loss of contrastive tone in Seoul and other dialects: the confusion and collapse of the contrastive system with the segment-induced system, which was regular. This explanation would hold regardless of the modern synchronic analysis.

While the change in progress in Seoul superficially suggests the emergence of contrastive tone, the evidence presented here suggests Seoul Korean has been tonal since the time of Middle Korean. Regardless, the initial neutralization or total merger of plain and aspirated segments would result in the reemergence of the contrastive function of the tones, and the partial removal of the contraints on the underlying representations. Whether this sort of change counts as tonogenesis is a matter of definition.

Chapter 9

Conclusion

9.1 Change in Progress in Seoul

I had an interesting conversation with speakers f21 and f22 after they finished their perception tasks. They told me that the two token types could be distinguished based on whether they were high or low, and f21 said that as the task went on she began to rely on pitch to determine her answers. At first I thought they were saying they were previously aware of this pitch distinction, which surprised me. However, they both said that they were not previously aware of the pitch distinction, but became aware of it during the course of the task.

This captures the gist of what the perception experiment was designed to show, that pitch has taken over the functional load of discriminating between plain and aspirated stops. That's certainly what the results of the experiment suggest: pitch extremes tend to override the role of VOT in discrimination, with higher pitches almost universally indicating /p^h/. I've suggested that reliance on the tonal contrast is driving the neutralization of the plain and aspirated series, by allowing speakers to relax the segmental contrast between the plain and aspirated stops.

The tonal contrast isn't fully explanatory however, since the same change hasn't occurred in other dialects, and presumably this change could have occurred much earlier if the tonal contrast goes back to Middle Korean. Seoul Korean is also different from other dialects in that it has relatively recently lost the vowel length distinction still present in other dialects. These changes may have been triggered by social forces, but I have no specific speculations on that.

The evidence from the 20 speakers reported on here suggests that the change in progress is not complete, simply because there is variation among even the youngest speakers. However, for reasons discussed in Chapter 5, it's not clear what the end point of the change, a true neutralization, would look like in terms of distributions. We simply need more data from more speakers to determine when the change has stabilized in the community, and how things will look at that point.

9.2 Phonological Implications

9.2.1 Tone

I suggest in Chapter 8 that Seoul Korean has underlying tones. One obvious advantage of this is that there's no need for some sort of phonological derivation of the surface tones based on segmental features. However, the disadvantage of my theory is that it doesn't recognize the synchronic regularity of the segment-tone correlation. Even loan words and nonsense words obey this correlation. My theory requires an active process of analogy, or a constraint on underlying forms, that prohibits the assignment of words to lexical classes based on their initial consonant. This, arguably, is theoretically objectionable. It's unclear how to distinguish between the two possibilities.

9.2.2 Plain and Aspirated Segments

Although the apparent time data from Silva (2006) is an important aspect of the paper, the focus of the paper is on the phonological implications of that data. Silva's analysis is designed to give a single underlying representation for all speakers in the community.

> As some younger speakers have neutralised VOT differences between lax and aspirated stop phonemes and mark this contrast tonally, formal accounts of the Korean obstruent system need revision in the direction of feature representations that ad- equately account for the phonetic behaviour of *all* speakers of the standard variety, young and old alike. The modification proposed here involves replacing underlying glottal aperture features (specifically [spread glottis] and [constricted glottis]) with a more abstract laryngeal 'tensity feature' (a' la Kim 1965), [stiff]. This single feature has the advantage of generalisability across the speaker pool: [stiff] may be phonetically implemented in either a more traditional way, i.e. maintaining VOT dis- tinctions between lax and aspirated stops, or a more innovative manner, i.e. backgrounding aspiration differences in favour of a tone-based strat- egy for marking the underlying lax vs. aspiration distinction. (Silva, 2006, p.288)

It's my view that it's not necessary for a single community to have a single underlying representation. What's more problematic for phonology here is that the change underway is gradual. Many different levels of VOT are possible phonetically, as seen just among the Korean speakers, and as pointed out in Cho and Ladefoged (1999). How much aspiration is necessary before we say that plain stops are [+aspirated] in initial position? If we maintain the underlying contrast of phonemes by, for example, claiming that plain stops are [+voice], we run into a problem in one of two different ways. We either have to assign two different VOTs to the same surface feature [+aspirated], or assign large VOTs to some feature other than [+aspirated] for the plain stops. For speakers with a true neutralization we avoid the problem, but the question is, does the community actually experience a discrete change from contrast to neutralization. My guess is that this is not the case, but more data from the intermediate ages in the community may show otherwise.

I also believe that features like [stiff] and [tense] are misplaced in representations of the aspirated series. These features took hold in the literature in part due to the supposed microprosodic effect of the H-segments, which is false, and in part due to the desire to create a natural class for the H-segments, which is not necessary if we allow for underlying tone/accent. Furthermore, the articulatory data from Hirose et al. (1974) show that laryngeal tension is clearly present in tense segments, but not apparent in aspirated segments.

9.2.3 Tense Segments

Recall the following:

> Kim has observed that the fundamental frequency in the vowel following release of a stop tends to be higher for the voiceless stops [p] and [p^h] than for the partially aspirated [p_k]. This finding supports the classification of the first two of these stops as [+stiff] and the last as [-stiff], since vocal-cord stiffness has an influence on the frequency of vibration. (Halle and Stevens, 1971, pp.206-207).

In this account, the Korean tense stops are the same as the cross-linguistically common unaspirated voiceless stops, and the Korean plain stops are a third, typologically unique class, which ironically are [-stiff]. This implies that the phonetic continuum of aspiration from none to parital to full would involve a change from [+stiff] to [-stiff] to [+stiff]. Had the authors made the tense stops typologically unique, the plain stops may have fit better into their overall system, perhaps by being phonologically unaspirated.

Tense stops are the main phonological problem presented by Korean, being otherwise unattested. In this case, I do favor a new feature, or a new combination of features, in the underlying representation. While it's possible to construct the underlying contrasts without an explicit [tense] feature, I think this unnecessarily divorces phonology from phonetics. If a feature is possible at the surface level, it seems unrealistic to say that it's not possible at the underlying level. Some authors analyze tense segments as geminate plain segments to represent the contrast without a new feature, but this results in the geminate having a property that the singleton does not. Recall the articulatory differences: plain stops involve glottal abduction, while tense stops are tightly adducted at the time of closure release.

Part of the argument for a geminate analysis of tense stops is their long length word-internally. Given the complex articulation of the glottis for tense stops, abduction followed by adduction during the course of the closure, we could analyze the segments as complex, rather than geminated plain stops. Historically, the initial tense stops originated from consonant clusters that in turn originated from syncope. What's particularly interesting about this is the fact that some authors believe that intervocalic voicing was also active in Middle Korean (Ramsey, 1991). Therefore syncope might have been bringing together an intial voiceless stop with an internal voiced stop. This is of course exactly what tense stops look like articulatorily, Kagaya (1974) remarking that the glottal position is similar to the position for voicing. Synchronically we could perhaps analyze tense stops as clusters.

9.2.4 Inter-sonorant Plain Segments

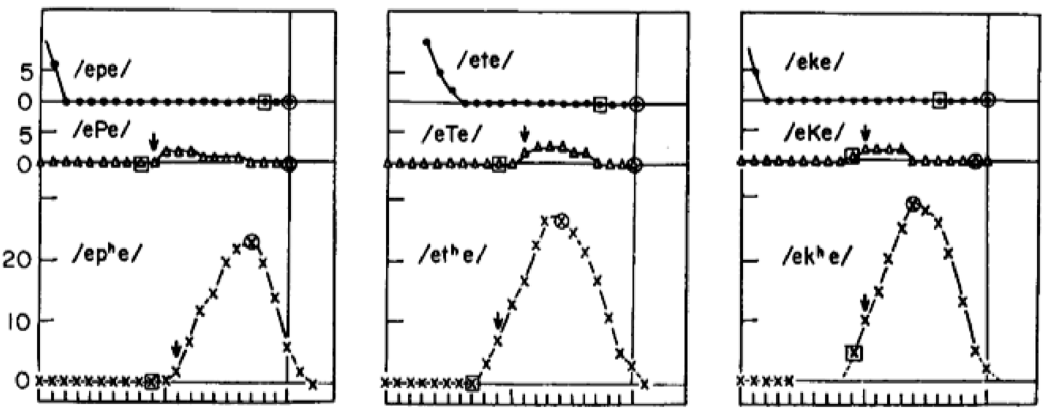
One question regarding plain segments is whether the inter-sonorant variant is a distinct voiced allophone, or whether it is passively voiced, a purely phonetic effect due to a weak prosodic position. Some authors suggest a purely phonetic explanation, for example:

> The variation of lenis stop voicing due to rate, phrasing, segmental and prosodic

contexts suggests that the lenis stop voicing rule in Korean is not a categorical phonological phenomenon but [a] gradient phonetic phenomenon. (Jun, 1996b, p.93)

The best counter evidence to this is the glottal width data from Kagaya (1974):

Figure 9.1: Glottal width data from Kagaya (1974)



The glottis is never abducted for intervocalic plain stops. It is true that this data set is limited and constrained, so let's also consider the results from Silva (1992) discussed in Chapter 2. There is a consistent result for all stop manners and prosodic positions of 14-20 ms for voicing into the closure, with the exception of intervocalic plain stops which have roughly double the voicing duration. This doesn't support a gradient analysis.

9.3 Future Work

The first task for continuing this line of research is an expansion of the data collection discussed in Chapter 4, and I plan to add many more speakers to that data set. The change underway in Seoul relates to many questions, the time course of the change, its end state phonetically, and its potential ramifications for the system. While I believe that lexical tone already exists, a neutralization of plain and aspirated stops is still just as interesting.

Will the AP-internal contrast maintain the distinction, or will a voiced series of phonemes emerge? Will words idiosyncratically change tonal classes now that the initial segment is no longer a reliable indicator?

More work can be done on the perceptual side of things as well. The experiment presented in Chapter 6 showed that pitch has a dominant effect, but wasn't precise beyond that. I plan to do a new perception experiment were the pitch of natural tokens is left unchanged, but the VOT is modified gradually. Removing pitch as a variable will allow for a finer examination of the role of VOT. Closure duration also deserves examination is this fashion; it's unclear if closure duration really serves a contrastive function here, or is simply an automatic phonetic effect.

Finally, I hope to expand the data collection to other dialects, even though a similar change is not underway elsewhere (to my knowledge). Of particular interest are the patterns of segment-correlated tone in dialects that maintain the contrastive tonal system of Middle Korean.

Bibliography

Ahn, S.-C., and G. K. Iverson. 2004. Dimensions in korean laryngeal phonology. *Journal of East Asian Linguistics* 13:345–379.

Bradshaw, M. M. 1999. A cross-linguistic study of consonant-tone interaction. Doctoral Dissertation, Ohio State University.

Cho, T., S. Jun, and P. Ladefoged. 2002. Acoustic and aerodynamic correlates of korean stops and fricatives. *Journal of Phonetics* 30:193–228.

Cho, T., and P. A. Keating. 2001. Articulatory and acoustic studies on domain-initial strengthening in korean. *Journal of Phonetics* 29:155–190.

Cho, T., and P. Ladefoged. 1999. Variation and universals in vot: evidence from 18 languages. *Journal of Phonetics* 27:207–229.

Choi, H. 2002. Acoustic cues for the korean stop contrast: dialectal variation. *ZAS papers in linguistics* 28:1–12.

Cole, J., H. Kim, H. Choi, and M. Hasegawa-Johnson. 2007. Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from radio news speech. *Journal of Phonetics* 35:180–209.

Cristo, A. Di, and D. J. Hirst. 1986. Modelling french micromelody: analysis and synthesis. *Phonetica* 43:11–30.

Halle, M., and K. Stevens. 1971. A note on laryngeal features. *MIT Research Laboratory of Electronics Quarterly Report* 101:198–213.

Han, M. S., and R. S. Weitzman. 1970. Acoustic features of korean /p,t,k/, /p,t,k/ and /ph,th,kh/. *Phonetica* 22:112–128.

Haudricourt, A. G. 1971. Two-way and three-way splitting of tonal systems in some far-eastern languages. In *Tai phonetics and phonology*, ed. J. G. Harris and R. B. Noss, 58–86. Central Institute of English Language, Mahidol University, Bangkok.

Hirose, H., C. Y. Lee, and T. Ushijima. 1974. Laryngeal control in korean stop production. *Journal of Phonetics* 2:145–152.

Hombert, J.-M. 1978. Consonant types, vowel quality, and tone. In *Tone: a linguistic survey*, ed. V. Fromkin, 77–111. New York: Academic Press.

Hombert, J.-M., J. J. Ohala, and W. G. Ewan. 1979. Phonetic explanation for the development of tones. *Language* 55:37–58.

Jessen, M. 1998. *Phonetics and phonology of tense and lax obstruents in german*, volume 44 of *Studies in fucntional and structural linguistics*. John Benjamins.

Jun, S. 1998. The accentual phrase in the korean prosodic hierarchy. *Phonology* 15:189–226.

Jun, S.-A. 1996a. Influence of microprosody on macroprosody: a case of phrase initial strengthening. *UCLA Working Papers in Phonetics* 92:97–116.

Jun, S.-A. 1996b. The phonetics and phonology of korean prosody. Doctoral Dissertation, Ohio State University.

Kagaya, R. 1974. A fiberscopic and acoustic study of the korean stops, affricates and fricatives. *Journal of Phonetics* 2:161–180.

Kang, O. 1992. Korean prosodic phonology. Doctoral Dissertation, University of Washington.

Kenstowicz, M., and C. Park. 2006. Laryngeal features and tone in kyungsang korean: a phonetic study. In *Studies in phonetics, phonology, and morphology (to appear)*.

Kim, C.-W. 1965. On the autonomy of the tensity feature in stop classification (with special reference to korean stops). *Word* 21:59–104.

Kim, C.-W. 1970. A theory of aspiration. *Phonetica* 21:107–116.

Kim, M.-R., and S. Duanmu. 2004. 'tense' and 'lax' stops in korean. *Journal of East Asian Linguistics* 13:59–104.

Kim, M.-R. C. 1994. Acoustic characteristics of korean stops and perception of english stop consonants. Doctoral Dissertation, University of Wisconsin-Madison.

Kim, N.-J. 1997. Tone, segments, and their interaction in north kyungsang korean: a correspondence theoretic account. Doctoral Dissertation, Ohio State University.

Kim-Renaud, Y.-K. 1974. Korean consonantal phonology. Doctoral Dissertation, University of Hawaii.

Kingston, John, and Randy L. Diehl. 1994. Phonetic knowledge. *Language* 70:419–454.

Ko, E.-S., N.-R. Han, Alexandra Canavan, and George Zipperlen. 2003a. Korean telelphone conversations speech. Linguistic Data Consortium, Philadelphia.

Ko, E.-S., N.-R. Han, Alexandra Canavan, and George Zipperlen. 2003b. Korean telelphone conversations transcripts. Linguistic Data Consortium, Philadelphia.

Labov, W. 1994. *Principles of linguistic change, volume 1: Internal factors*. Blackwell.

Lisker, L. 1957. Closure duration and the intervocalic voiced-voiceless distinction in english. *Language* 33:42–49.

Lisker, L., and A. S. Abramson. 1964. A cross-language study of voicing in initial stops; acoustic measurements. *Word* 20:384–422.

Maddieson, I. 1997. Phonetic universals. In *The handbook of phonetic sciences*, ed. J. Laver and W. J. Hardcastle, 619–639. Blackwell.

Nearey, T. M. 1980. On the physical interpretation of vowel quality: cinefluorographic and acoustic evidence. *Journal of Phonetics* 8:213–241.

Ramsey, S. R. 1975. Accent and morphology in korean dialects: a descriptive and historical study. Doctoral Dissertation, Yale University.

Ramsey, S. R. 1991. Proto-korean and the origin of korean accent. In *Studies in the historical phonology of asian languages*, ed. W. G. Boltz and M. C. Shapiro, volume 77 of *Current Issues in Linguistic Theory*, 215–238. John Benjamins.

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

Shen, Z. W., C. Wooters, and W. S.-Y. Wang. 1987. Closure duration in the classification of stops: A statistical analysis. *Ohio State University Working Papers in Linguistics* .

Silva, D. J. 1992. The phonetics and phonology of stop lenition in korean. Doctoral Dissertation, Cornell University.

Silva, D. J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary korean. *Phonology* 23:287–308.

Silverman, K. 1986. F0 segmental cues depend on intonation: the case of the rise after voiced stops. *Phonetica* 43:76–91.

Stevens, K. 1998. *Acoustic phonetics*. MIT Press.