The Phonetics of VOT and Tone Interaction in Cantonese

By Holman B. Tse

THESIS
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR A DEGREE OF
MASTER OF ARTS
IN THE
DEPARTMENT OF LINGUISTICS
UNIVERSITY OF CHICAGO
CHICAGO, IL
AUTUMN 2005

The Phonetics of VOT and Tone Interaction in Cantonese

Abstract

This study investigates the possible effects of lexical tone on Voice Onset Time (VOT) in Cantonese, a tonal language with a two-way contrast between short-lag (voiceless unaspirated) and long-lag (voiceless aspirated) stops. VOT was measured as the time interval between the stop burst and the onset of voicing for the following vowel. The recorded speech of 6 native speakers each producing 10 repetitions of 20 different words contrasting in aspiration and tone was analyzed. Tokens from each individual subject were divided into two sets for the purpose of comparison. The first set involved a comparison between the effects of a high-level 55 tone and a mid-level 33 tone. Results showed no significant VOT differences unless aspirated and unaspirated stops were examined separately. In this case, only the aspirated stops showed a significant difference with the 33 tone associated with higher VOT. The second set of stimuli compared the effects of 4 different phonemic tone categories (55, 25, 33, and 21) on VOT. Results show that words beginning with a lower tonal onset (and thus the 25 and 21 tones) correlated with higher VOT than words beginning with a higher tonal onset (the 55 and 33 tones).

1. Introduction

Ever since Lisker and Abramson (1964) introduced the concept of Voice Onset Time (VOT), many linguists have sought to test the utility of this phonetic concept as a timing cue used to distinguish different phonological categories. On the most basic level, VOT refers to the time interval between the release of a stop and the onset of voicing for the following vowel. Voiced stops by definition have negative VOT while voiceless unaspirated stops have relatively low positive VOT values (short-lag) and voiceless aspirated stops have relatively high positive VOT values (long-lag). While the VOT concept is most useful for distinguishing stops that contrast in either voicing or aspiration, the VOT values for any given phoneme are not fixed. They can vary in response to prosodic context, rate of speech, place of articulation, vowels, and other variables. For this reason, studies that examine the stability of VOT across different phonetic environments provide an interesting look at the phonetics-phonology interface.

Building on previous research examining the different variables that can affect VOT, this thesis aims to explore the possible effects of one variable that has received little attention in the literature: lexical tone. Does tone have any effect on VOT? For example, could a syllable with a high-level tone have an onset stop with a higher or lower VOT value than an onset stop of a syllable with a low-level tone? If tone does turn out to have consistent effects on VOT values, what would be the implications of these effects on what a number of linguists have described as the “flexible” (Repp and Libermann 1987) nature of phonetic categories in general?

To explore such questions, this paper will focus on Cantonese as a case study. In particular, the generalization made by Hyman and Schuh (1974) and refuted by Maddieson (1974, 1976, and 1978) that “consonants affect tone, but tone does not affect consonants” will be evaluated. The paper will begin with a review of the relevant literature on VOT and tone in Sections 2 and 3. Section 4 includes general information about the tone and stop inventory of Cantonese. This will be followed by a presentation of an experiment run to address the question of tone and VOT interactions in Section 5 with the results presented in Section 6. Section 7 discusses the results. Section 8 is a conclusion with a brief discussion of future research possibilities.

2. The Implications of VOT on Phonological Categorization

2.1 VOT and Linguistic Typology

In the original study that coined the term ‘Voice Onset Time’, Lisker and Abramson (1964) present measurement data from 11 languages including Cantonese. They found that in spite of variation in the precise VOT values for different stops in these languages, VOT values still managed to cluster into a maximum of three groupings, depending on the language, that correspond to the different types of stop phonemes that exist in a given language. These three groupings include voiced stops, which all have negative VOT values, voiceless unaspirated stops, which all have positive VOT values that cluster around zero, and voiceless aspirated stops, which have positive VOT values in a range that is consistently distinct for the range found for unaspirated stops in languages that have an aspiration contrast. Based on the clustering of VOT values, Lisker and Abramson defined languages as two-category, three-category, or four-category.

Two-category languages either have a contrast between voiced and voiceless unaspirated stops as in Dutch, Spanish, Hungarian, and Tamil or have a contrast between voiceless unaspirated and voiceless aspirated stops as in Cantonese and English. Three-category languages include Eastern Armenian, Thai, and Korean. While the inventories of the former two consist of voiced, voiceless unaspirated, and voiceless aspirated stops, the Korean categories include one described as tense, long, and glottalized (with positive VOT values near zero), another described as lax and slightly aspirated (with VOT slightly higher than tense stops), and a third described as lax and heavily aspirated (with VOT higher than the other two categories). All three categories are voiceless and, thus, have positive VOT. Though Korean appears unique among three-category languages in not having its three stop categories more spread out along the VOT continuum, it is quite clear that different ranges of VOT values still manage to distinguish these three categories. Finally, the four-category languages include Hindi and Marathi, which have the same three stops found in Thai and Armenian in addition to voiced aspirated stops. Because of overlap in VOT values for voiced aspirated and unaspirated stops, Lisker and Abramson conclude that VOT is not useful for distinguishing all the stops in these languages.

While Lisker and Abramson (1964)'s study focused on phonetic details, it left many questions about the phonological implementation of these details unanswered. Many subsequent studies explored how VOT can be incorporated into various phonological models of distinctive features. Keating (1984), for example, points out that phonetic categories need not map one-to-one onto phonological categories. For instance, in Halle and Stevens (1971)'s model, features that describe stop contrasts include [±stiff vocal cords], [±slack vocal cords], [±constricted vocal cords], [±spread vocal cords]. This is unnecessarily complicated because it takes into account too many phonetic details, many of which do not seem to be phonologically important in any language. For example, this feature system distinguishes between voiced laryngealized stops from true implosives and voiceless unaspirated from voiceless lax but no known language in the world makes such contrasts phonological. As Keating states:

> "A set of features predicated on phonetic accuracy will require ever more additional features as new articulatory mechanisms are discovered. The proliferation of features is the price paid for using the same set of features for phonological as well as low-level phonetic representation – an otherwise appealing constraint on the relation of phonology to phonetics. … H&S [Halle and Stevens 1971] and (SPE) don't simply have the wrong features … they will ALWAYS have TOO MANY [emphasis in original] features because they want to describe exactly how individual sounds are articulated. While we want the phonological features to have some phonetic basis, we also want to distinguish possible contrasts from possible differences. Our goal must be to find some feature framework in which the phonetic basis of phonological features is not explicit phonetic detail." (1984: 289)

Taking Polish and English as examples, both are two-category languages with voiceless unaspirated stops. While the second category in Polish is pre-voiced, the second category in English is aspirated. Keating argues that a polarization principle operates in these two languages such that in English, the voiceless unaspirated stop is slightly pre-voiced while in Polish this sound is slightly aspirated. This polarization principle maximizes the phonetic differences in language specific stop categories. Yet, in spite of the different phonetic sounds that are actually produced in these languages, the universal feature [±voice] is all that is needed in phonological representations of these languages even if, as is the case for English, the voiceless unaspirated stop has both voiced and voiceless allophones. [+voice] will cover both allophonic variants.

2.2 Factors that affect VOT

While the previous section makes clear that VOT values cluster together in an organized way to maintain voicing and aspiration contrasts, the picture gets more complicated when we examine some of the wide range of factors that can shift VOT values in different directions without losing these contrasts. One of these factors, place of articulation, has been widely attested cross-linguistically. Even Lisker and Abramson (1964) noticed a pattern with the place of articulation in their data but they did not try to account for why place of articulation may affect

VOT. In a study of 18 languages, Cho and Ladefoged (1999) show that the further back the place of articulation, the longer the VOT for voiceless stops especially for velars and uvulars compared to coronals and labials although this generalization holds less true for the difference between labials and alveolars and the difference between velars and uvulars. Cho and Ladefoged suggest that some of the deviations from this place of articulation generalization may be due to language-specific effects interacting with universal phonetic constraints. For instance, some of the languages they studied contrast plain oral and ejective stops. In this case, shifting VOT values can be seen as a way of maximizing contrasts between different stop categories.

Another factor shown to affect VOT values is prosodic position. In English, for example, stress enhances VOT contrasts such that in stressed syllables pre-vocalic voiceless stops have higher VOT while underlying ‘voiced’ stops are more likely to be pre-voiced than in other prosodic positions (Lisker and Abramson 1967). Increasing the VOT differences between the two stop categories can, thus, be seen as a way of using stress to emphasize this distinction. Van Dam (2003) shows other stress-induced patterns in English. First of all, he shows that underlying voiceless stop onsets in stressed syllables have higher VOT than onset stops in syllables following stress. This study also shows that onset stops in a position not adjacent to stressed syllables have intermediate VOT values. These prosodic effects are language-specific as Dutch has been shown to display the opposite effect with underlying voiceless stops having shorter VOT in stressed contexts (Cho 2003). Cho attributes the difference between English and Dutch to the different phoneme inventories of the two languages. Yet, the same principle of prosodic strengthening operates in both languages. While Dutch has a contrast between voiced and short-lag voiceless stops, the two stop categories in English are both voiceless. The difference can be seen in the fact that in Dutch, the voiceless stop involves glottal adduction while in English the voiceless stop involves glottal abduction. Prosodic strengthening in Dutch would mean more adduction and hence lower VOT while prosodic strengthening in English would mean more abduction and hence higher VOT. In all of these cases, the prosodic effects on VOT values can be seen as a way of maintaining the phonemic contrasts that exist in a particular language by enhancing the differences.

A third variable that has been shown to affect VOT values is rate of speech. Kessinger and Blumstein (1997) examine the effects of speaking rate in three different languages with different stop inventory systems: Thai (pre-voiced, short-lag, and long-lag), French (pre-voiced and short-lag), and English (short-lag and long-lag). As rate of speech increases, VOT values for both voiceless aspirated and voiced stops in the languages that have them approach 0. These sounds, thus, become pronounced with shorter duration. A slower speaking rate results with the inverse of these effects and, thus, more negative VOT for pre-voiced stops and more positive VOT for long-lag stops. Interestingly, the VOT for short-lag stops remains relatively stable at all rates of speech. There was also no overlap found between pre-voiced and short-lag and only minimal overlap found between short-lag and long-lag at faster rates of speech. The short-lag category, which is found in all three languages studied, seems to act as a “phonetic anchor” to help maintain stop contrasts by keeping the VOT values of one category in place. While Kessinger and Blumstein’s study investigated speech production, Summerfield (1981) and Boucher (2002) investigate speech perception of VOT boundaries. Both of these studies show that listeners learn to compensate for speaking rate effects by recognizing intrinsic timing ratios. These studies show that in spite of variation in absolute VOT values across speaking rates, stable boundaries that separate stop categories can also be found when intrinsic timing relations such as the ratio of VOT/syllable duration are examined.

Finally, to conclude this section, a few other factors affecting VOT worth mentioning briefly include vowels and sonorants. Stops preceding high vowels and sonorants have higher VOT than stops preceding mid and low vowels (Klatt 1975).

3. Tone and Consonant Interactions

The literature on the various factors that can affect VOT raises the question of what else can change VOT. This leads to the main topic of this paper, which is an investigation of lexical tone as a possible factor. Could tone have any effect on changing VOT values like rate of speech, vowel height, stress or other factors can?

Exploring the interaction between VOT and tone would be interesting because tone, like VOT is also a concept that is not defined by absolute boundaries or range of values. Like the distinction between aspirated and unaspirated stops, the distinction between different tone categories is defined in terms of contextually-dependent variables. Yip (2002) describes the movement from the purely phonetic to the phonological as the movement from fundamental frequency to pitch to tone. The fundamental frequency, or F0, is an acoustic term referring purely to the production of the sound itself in terms of the absolute number of pulses per unit of time. Pitch, on the other hand, is a perceptual term with qualitative distinctions based on relative factors. So for instance, the difference between a high and a low pitch sound for two different speakers may be based on significantly different F0 values. The F0 value for a given level of pitch can also vary contextually for an individual speaker such that a listener can unconsciously compensate for F0 differences and not perceive a pitch difference even with significantly different F0 values. Tone refers to pitch differences that take on a phonemic status. In other words, when two words differing only in pitch differences form a minimal pair, the pitch differences are considered tone differences.

The question of tone and VOT is also interesting because it is fundamentally an issue involving the interaction of tone and consonants. This is an area that has spawned much research especially in relation to tonogenesis, the historical development of tone. This research, however, has often assumed a single direction of effects that runs counter to the main question of investigation of this paper. For instance, Hyman and Schuh (1974) make the generalization that "consonants affect tone, but tone does not affect consonants." If this generalization is correct, it seems to suggest that tonal effects on VOT do not exist. Maddieson (1974, 1976, and 1978), however, disagrees with this generalization and cites numerous examples that he claims are exceptions that need to be reinterpreted.

One example is Jingpho, which exhibits a voicing alternation between [yàk] ('difficult') and [yàggai] ('it is difficult'). This alternation does not occur in words associated with high tone. For example, there is no voicing alternation between [cát] ('tight') and [cáttai] ('it is tight'). Maddieson (1974) argues that this is a counterexample to Hyman and Schuh's (1974) generalization because it clearly shows a case of low tone triggering a change on the following consonant. Hyman (1976) responds by saying that the Jingpho case is not an exception because it involves breathiness and it happens to be the case that all Jingpho low tones are accompanied by breathiness. He also reclarifies his original claim by restating it as "consonants affect pitch, but pitch does not affect consonants" (1976: 94). This reformulated statement excludes cases such as the Jingpho example, which he believes involve segmental changes triggered by breathiness rather than by pitch.

Hyman also clarifies his original statement by saying that "what we are interested in is cases where a tone possibly affects the <u>segmental</u> [emphasis in original] characteristics of a consonant" (1976: 92). Thus, relatively minor phonetic changes such as VOT changes induced by tone would be excluded from Hyman and Schuh's (1974) generalization. Nevertheless, Maddieson (1978) pushes the issue of directionality of effect even further by citing some examples from instrumental studies that show tone affecting consonants without changing segmental characteristics. For example, Maddieson (1978) makes the generalization that "higher tones condition longer consonants" (1978: 103). This claim is based on several sources including his own work on Thai which showed that the voiced bilabial stop /b/ has a longer duration when preceding the high and falling tones than when preceding the low, mid, and rising tones. He also cites Zee (1977) which shows that the duration of /s/ in /si:/ in Taiwanese is longer when preceding high tones. Gandour and Maddieson (1976) provide further evidence showing that tones can affect consonants. This study involved measurements of larynx movement for a Thai speaker and showed that when producing low tones, the larynx moved more downward than when producing high tones. This physiological effect shows that the following tone can affect a preceding consonant.

Though Maddieson (1974, 1976, and 1978) presents many examples that he believes show tone affecting consonants, he still agrees with Hyman (1976) that clear cases of consonants affecting tone rather than the reverse are far more common. Among the best cited cases of consonants affecting tone are accounts of tonogenesis from the loss of consonantal contrasts. Haudricourt's (1954) pioneering study showed that prevocalic consonants directly influence the F0 onset of the following vowel and that postvocalic consonants are crucial in determining the overall tone contour in a given syllable. Hombert's (1975 and 1978) studies involved detailed instrumental studies of the differential effects on F0 onsets of different consonant types and speculated about how these phonetic studies can shed light about the historical development of tone. The most widely agreed on effect is the development of high and low tones from consonantal voicing contrasts. Voiceless stops have been shown to raise the F0 onset while voiced stops lower F0 (Hombert 1975 and 1978). For example, if a language has a contrast between high and low tones such as in [tà]/[tá], it is often the case that this contrast developed from the minimal pair [da]/[ta] in an earlier stage of the language with the low tone corresponding to a word beginning with the voiced stop [d] and a high tone corresponding to a word beginning with the voiceless stop [t] (Yip 2002: 37).

While there seems to be consensus on the differential effects of voiced and voiceless stops on F0 onsets, there is less agreement about whether or not voiceless unaspirated and aspirated stops should show different effects on F0. Zee (1980) shows that in Cantonese, voiceless aspirated stops raise the F0 onset of the following vowel. Xu and Xu (2003), on the other hand, show the opposite effects for Mandarin. Hombert and Ladefoged (1977) show no significant difference between the effects of voiceless aspirated vs. unaspirated stops on the following F0 in English and French, which are both non-tonal languages. Yet, even different studies of the same language have shown contradicting results. For example, while Gandour (1974) shows that in Thai the F0 onset is lower following voiceless aspirated stops than following voiceless unaspirated stops, Ewan (1976) shows higher F0 following aspirated stops.

Lai (2004) speculates that the lack of agreement in these studies is due to the fact that they did not take into account factors such as place of articulation and following vowels. In designing her study on the effects of aspiration on F0, Lai (2004) took into account these factors and also took measurements of F0 onset, F0 at regular intervals, end F0, tonal duration, and VOT, which is something lacking in many studies of consonant and tone interaction. Results from this study showed that aspiration has a raising effect on F0 onset and that this raising effect disappears around the midpoint of the tonal duration. Though VOT was not the main focus of this paper, Lai does mention that results from a correlation test show that higher VOT values tend to co-occur with lower F0 onsets. Her explanation of this is that while aspiration basically raises F0 due to faster airflow, the effect decreases over time as aspiration gets longer and as airflow decreases with time. Thus, longer VOT translates into a disappearance of the F0 raising effect that would otherwise occur with aspiration.

From this section, it should be clear that the interaction of tone and consonants is a complex phenomenon that involves looking at many factors and multiple possibilities. It is an active area of research that has much to contribute to our understanding of the mechanisms of speech. The purpose of the current project described in the following pages is to contribute to this literature by applying the VOT concept as a tool for investigating tone and consonant interaction. This study also addresses whether or not tone affects VOT values and by doing so this study responds to Maddieson's (1974, 1976, and 1978) claim that it is possible for tone to affect consonants. While many of the other studies cited in this section were designed to test consonantal effects on F0 or tone, this study is designed to test the reverse possibility.

4. Cantonese Tones and Stops

Cantonese is a two-category tonal language that contrasts between short-lag and long-lag voiceless stops. The stop inventory includes pairs of bilabial, alveolar, and velar stops. One pair of alveolar affricates involving contrastive aspiration is also part of the consonantal inventory. All of these sounds are represented in IPA in Table 4.1.1.

Table 4.1.1: Cantonese Sounds Using Contrastive Aspiration

	Bilabial	Alveolar	Velar	Affricates
Aspirated	pʰ	tʰ	kʰ	t͡sʰ
Unaspirated	p	t	k	t͡s

The tonal inventory of the language includes six contrasting tones. In Table 4.1.2 and in the rest of the paper, the Chao (1930) tone numbering system in which relative pitch is represented on a scale of 1-5 with 5 being the highest pitch will be used. 'q' represents a checked tone, which occurs only in closed syllables. The checked tones 5q, 3q, and 2q occur in complementary distribution with the 55, 33, and 22 tones respectively and are essentially shortened versions of this latter group of tones. Though many linguists would not treat the checked tones as separate tones, I have included them in the table anyway to show the traditional tonal classification scheme.

Table 4.1.2: Cantonese Tonal Inventory Chart (Traditional Classification)[1]

	Sample Word CV(C) + Chao tone # (including alternate representations)	Tone Description	Chinese Character	Gloss
Yin (High Register)	si 55/53[2], 44[3]	High-level	詩	‘poem
	si 25, 35	Mid(high)-rising	史	‘history’
	si 33, 44	Mid-level	試	‘to try’
	sik 5q	High	識	‘to know’
	sɛk 3q, 4q	Mid	錫	‘to kiss’
Yang (Low Register)	si 21, 11, 22	Low-falling	時	‘time’
	si 23, 13, 24	Low-rising	市	‘city’
	si 22, 33	Low-level	事	‘matter’
	sik 2q	Low	食	‘to eat’

Though there are words pronounced as /si/ with all six phonemic tones, this is not the case for every possible segment sequence in Cantonese. Not all possible onset consonants are associated with actual words for all six tones. For instance, there are no words with the 22 tone that begin with aspirated stops. There are also no words beginning with unaspirated stops associated with the 21 and 23 tones. The few exceptions include onomatopoeic words and words formed through lexical reduplication. Hashimoto (1972: 110) briefly mentions these lexical gaps, but fails to give any explanation as to why they exist. In any case, these lexical gaps mean that it

---

[1] Information in this table comes from several sources including Hashimoto (1972), Yip (2002), Chen (2003), and Matthews and Yip (1994), which does a good job at presenting some of the variations in representation.

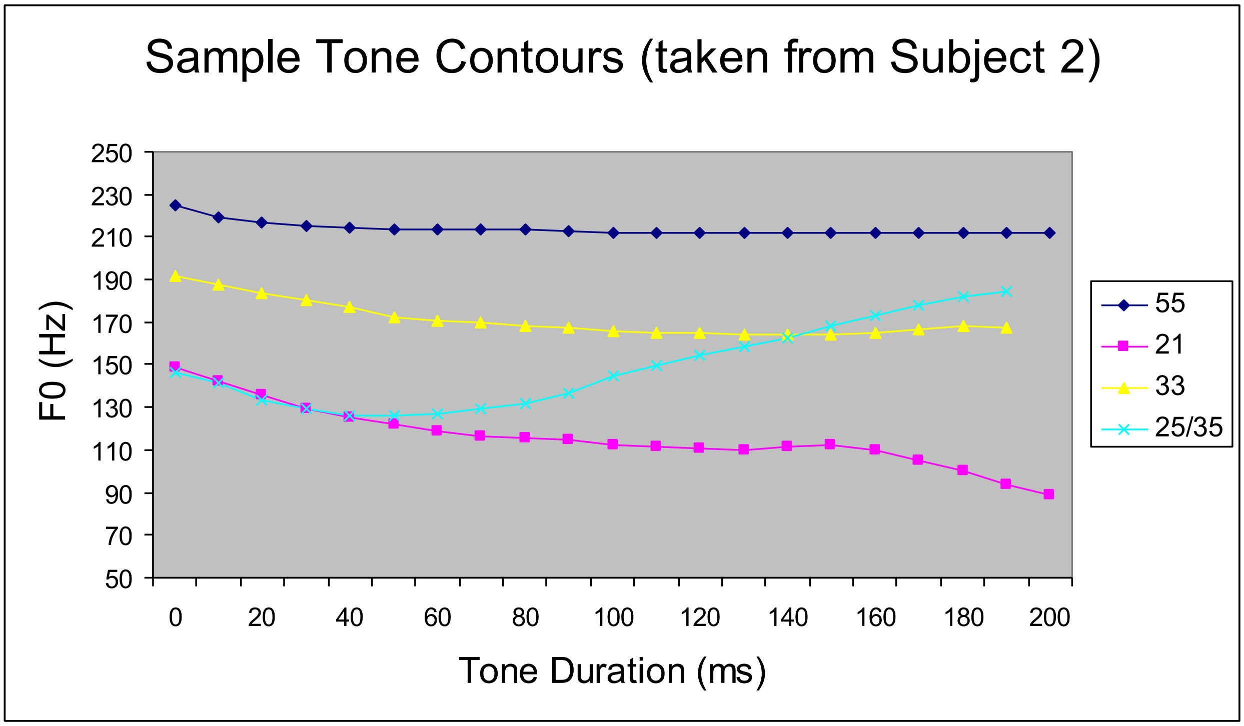
[2] The 55 and 53 tones have merged for most Cantonese speakers.

[3] Commas (,) that appear in this column are used to separate different representations of each tone category that have appeared in the literature. The Chao number representation for Cantonese tones that I adopt in the rest of this paper includes the first entry that appears for each category in this chart. Thus, they are 55, 25, 33, 21, 23, and 22 for the non-checked tones.

is not possible to make direct comparisons between aspirated and unaspirated stops for all 6 tones. Stimuli need to be selected carefully in order to make this study as controlled as possible.

With regards to tone, it is also worth noting that Cantonese has been described as a sandhi-poor language with the few tonal changes better described as "tone change" due to morphological processes rather than "tone sandhi" due to phonological processes (Chen 2003, Matthews and Yip 1994). One exception to this is the case of the 53 tone, for speakers who produce it, becoming a 55 when preceding another high-level tone. Yip (2002: 175-176) presents a phonological analysis of this process. In any case, the overall lack of tone sandhi means one less factor to worry about in designing this study.

Finally, Chart 4.1.3 is a graphic representation of the tonal contour inventory of one of the subjects. The chart includes only the four tone categories examined in this study. It is generally representative of all the subjects who participated in this study and suggests that the tones should be classified as 55, 33, 25, and 21.



5. Methods/Procedures

5.1 *Stimuli*

Table 5.1.1: Part 1 – Place of Articulation/High vs. Mid Tones

Tone	pʰou	pou	tʰou	tou	t͡sʰou	t͡sou
55	鋪床 'to make bed'	煲水 'to boil water'	滔滔不斷, ‘overflow of water’	刀片'razor blade'	粗口 ‘vulgar, rough sounding’	租房 ‘to rent, lease’
33	鋪頭 'a store, shop'	報紙 'newspaper'	吐痰 ‘to spit phlegm’	到達 'arrive, reach’	醋米 ‘rice vinegar’	灶頭 ‘stove top’

Table 5.1.2: Part 2 – Different Lexical Tones

Tone	pʰa	pa
55	趴地‘lie down on the floor’	巴士 ‘bus’
25/35	(豬)扒飯 ‘pork chop with rice’	把手 ‘handle’
33	怕怕 ‘scared’	霸位 ‘to hog seats’
21	爬山 ‘climb mtn’	爸爸 ‘daddy’

As can be seen in Tables 5.1.1 and 5.1.2, stimuli were divided into two groups. In Part 1, the stimuli all have a rime of /-ou/ and have onsets that contrast in aspiration across two places of articulation and across two tonal categories. Other than the bilabial and alveolar stops, affricates were also included. In Part 2, a wider range of tonal categories was examined. All stimuli have the rime /-a/ and the same place of articulation, so these variables were not a concern. Even though Cantonese has a total of six contrastive tonal categories, not all of them are included in this study because of lexical inventory gaps. There was also a preference for keeping this study as controlled as possible and this consequently meant sacrificing the inclusion of two of the six tone categories due to complications created by these inventory gaps.

5.2 Subjects

Six subjects (five male and one female), all native Cantonese speakers from Hong Kong in their early 20's, were paid to participate in this study. All have lived in the United States for less than five years and are University of Chicago students. None of them reported any speech or hearing problems. Recordings of their speech were made using a solid state recorder in a sound proof booth in the Phonetics Laboratory of the University of Chicago.

5.3 Procedures

Each subject was asked to read from a prepared list that includes the 20 words in Tables 5a and 5b appearing 10 times each in a randomized order and placed in a carrier phrase. A computer program automatically generated the list allowing each subject to read the phrase from a screen and to click the mouse when s/he was ready for the next word. Since written vernacular Cantonese is not a "standard" language in the socio-political sense, it was important to have each subject read through the list before recording to make sure s/he knew how to pronounce each word correctly with the right tone. For cases where ambiguity in meaning or pronunciation could potentially arise and for consistency, the target word was put in a contextual phrase XY so that it would be easier for subjects to correctly identify each word. For example, the character 鋪 (pʰou) could either be pronounced with a 55 or with a 33 tone. The character that follows, 床 (t͡sʰɔŋ33) or 頭 (tʰɐu25), disambiguates the meaning and pronunciation of 鋪 (pʰou). Thus, 鋪床 (pʰou55 t͡sʰɔŋ33) is pronounced with a 55 and means 'to make the bed' while 鋪頭 (pʰou33 tʰau25) is pronounced with a 33 and is a noun referring to 'a shop' or 'a store'. The contextual phrase always preceded the following carrier phrase with the first syllable in the contextual phrase (this was always the target word) inserted into the carrier sentence:

(1) XY, 我會讀 Y 俾你聽
XY, ŋɔ23 wui23 tuk22 X pei25 nei23 tʰɛŋ55[4]
'XY, I will read X to you.'

Here is an example of a line of text that subjects were expected to read:

(2) 鋪床, 我會讀 鋪 俾你聽
pʰou55 t͡sʰɔŋ33, ŋɔ23 wui23 tuk22 pʰou55 pei25 nei23 tʰɛŋ55
XY, I will read X to you. (X = pʰou55)

The first element underlined is the target word placed in a contextual phrase XY. The target word, X, is then placed into the carrier phrase with the same tone as in the context phrase. A total of 200 tokens were recorded for each speaker adding up to a grand total of 1200 tokens examined in this study.

5.4 Measurements

After completing the recordings, the next step was to analyze the recordings in digitized form using the software PRAAT (Boersma and Weenink 2005). Analysis first involved segmenting all 1200 tokens collected for this study. An example of segmentation to determine the VOT boundaries is found in Figure 5.4.1 below.

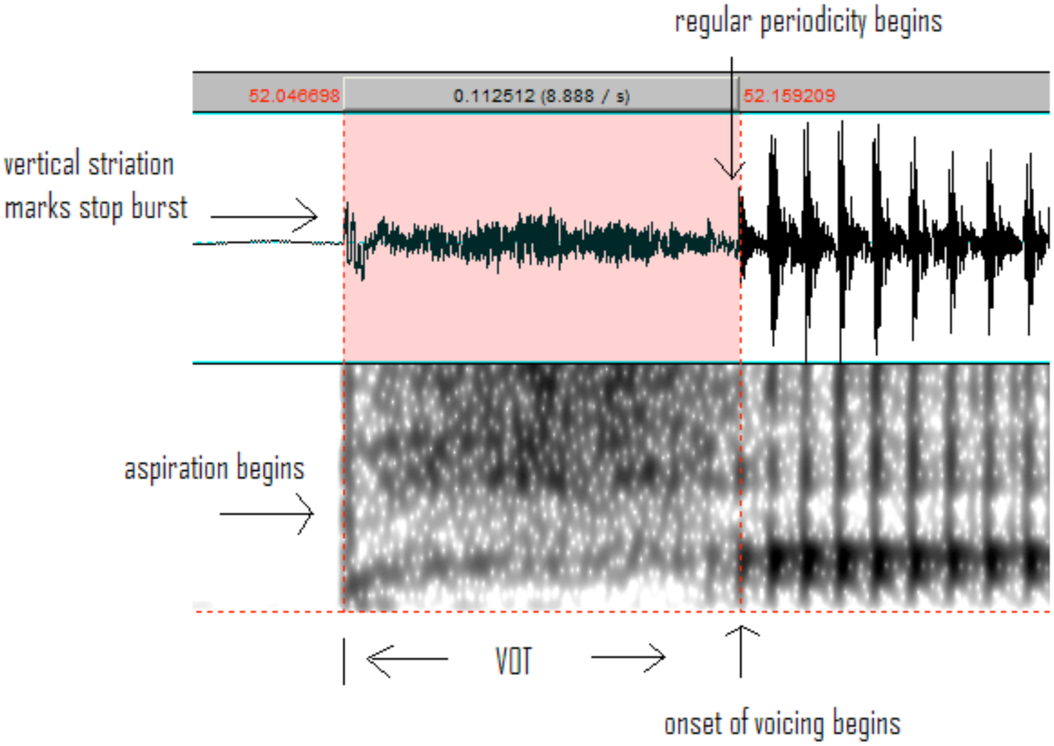
Sudden vertical bursts in the waveform were generally used to mark the boundary for the beginning of both aspirated and unaspirated stops. The end of VOT was determined by finding the point on the time-axis where periodicity clearly begins. The same procedure was used for affricates as well. The only difference is that the VOT for affricates also includes a period of aperiodic noise associated with the /s/ part of the sound. In addition to VOT, vowel length was also measured so that word length could be determined from the sum of VOT and vowel length.

---

[4] The phonetic pronunciation spelled out here is actually a conservative pronunciation of the phrase. In recent years, Hong Kong Cantonese has experienced a series of sound changes including the loss of the velar nasal onset, /ŋ/ > Ø or /ŋ/ > /ʔ/ for some speakers, the merger of the alveolar nasal and liquid /n/ > /l/ onsets, and alveolarization of codas, /ŋ/ > /n/ and /k/ > /t/ (Zee 1999). All six subjects participated in these sound changes to varying degrees. Subjects who were the most advanced in these changes would have pronounced the carrier phrase as ʔɔ23 wui23 tut22 __ pei25 lei23 tʰɛn55.

Following segmentation of both VOT and vowel length, a script was run on the segmented files that automatically calculated the length of the VOT and vowel length boundaries defined for each token. For the final stage of analysis, the ratio of VOT/Word length was used rather than absolute VOT values so that the effects of speaking rate can be controlled. For Part 1, a three-way ANOVA was performed to test the effects of tone, aspiration, and place of articulation. Part 2 involved a two-way ANOVA that tested only the effects of tone and aspiration.

Figure 5.4.1: An Example of Segmentation on PRAAT



6. Results

6.1 Part 1 Results

Tables 6.1.1, 6.1.2, and 6.1.3 show the complete results for Part 1. These tables include the average and standard deviation of the VOT/Word Length ratio of all Part 1 tokens. The results are organized such that it is possible to look at the data in several different ways. For instance, one could look at the results for all tone 55 words, for all labial aspirated stops, for all affricates, etc.

Table 6.1.1: Average VOT/Word Length for Aspirated Stops

Place	Tone	Average	Std. Deviation	N
Labial	55	0.2713	0.07365	60
Labial	33	0.2949	0.05902	60
All Labials		**0.2831**	**0.06751**	**120**
Alveolar	55	0.2812	0.07070	60
Alveolar	33	0.2862	0.06404	60
All Alveolars		**0.2837**	**0.06722**	**120**
Affricate[5]	55	0.3977	0.06768	60
Affricate	33	0.4106	0.06768	60
All Affricates		**0.4041**	**0.06405**	**120**
All Places	**55**	**0.3167**	**0.08910**	**180**
All Places	**33**	**0.3306**	**0.08727**	**180**
All Aspirated		**0.3236**	**0.08727**	**360**

Table 6.1.2: Average VOT/Word Length for Unaspirated Stops

Place	Tone	Average	Std. Deviation	N
Labial	55	0.0701	0.03788	60
Labial	33	0.0667	0.03201	60
All Labials		**0.0684**	**0.03496**	**120**
Alveolars	55	0.0592	0.02389	60
Alveolars	33	0.0595	0.02201	60
All Alveolars		**0.0593**	**0.02287**	**120**
Affricates	55	0.2355	0.08136	60
Affricates	33	0.2349	0.07993	60
All Affricates		**0.2352**	**0.08031**	**120**
All Places	**55**	**0.1216**	**0.09687**	**180**
All Places	**33**	**0.1204**	**0.09629**	**180**
All Unaspirated		**0.1210**	**0.09629**	**360**

---

[5] Though the place of articulation for all affricates in Cantonese is technically alveolar, affricates were treated as being part of a separate place of articulation because the extra frication associated with these sounds makes the VOT of affricates significantly longer than for other alveolar segments.

Table 6.1.3: Average VOT/Word Length for Both Aspirated and Unaspirated Stops

Place	Tone	Average	Std. Deviation	N
Labial	55	0.1707	0.11662	120
Labial	33	0.1808	0.12398	120
All Labials		**0.1757**	**0.12021**	**240**
Alveolar	55	0.1702	0.12324	120
Alveolar	33	0.1729	0.12340	120
All Alveolars		**0.1715**	**0.12307**	**240**
Affricates	55	0.3166	0.10819	120
Affricates	33	0.3227	0.11495	120
All Affricates		**0.3197**	**0.11143**	**240**
All Places	**55**	**0.2191**	**0.13484**	**360**
All Places	**33**	**0.2255**	**0.13884**	**360**
All Tokens		**0.2223**	**0.13680**	**720**

Table 6.1.4 summarizes the results of the three-way ANOVA performed on the Part 1 data. As shown here, both aspiration and place were shown to have significant effects on VOT. Aspiration should not be a surprise since the contrast between aspirated and unaspirated stops is a phonemic contrast in Cantonese. The difference in VOT between these two categories should be significant as studies such as Kessinger and Blumstein (1997) show little overlap between the VOT values of short-lag and long-lag stops even across different speaking rates. Place of articulation was also shown to be significant, but a closer look at the pairwise comparisons shows that this is due to the inclusion of affricates.

Table 6.1.4: Significance of Factors on VOT/Word Length

Factors	F =	p =	Significance
Aspiration	(1, 708) = 2077	< 0.01	*
Place	(2, 708) = 479.5	< 0.01	*
Tone	(1, 708) = 2.024	0.155	n.s.
Aspiration*Place	(2, 708) = 14.78	< 0.01	*
Aspiration*Tone	(1, 708) = 2.871	0.091	n.s.
Place*Tone	(2, 708) = 0.232	0.793	n.s.
Aspiration*Tone*Place	(2, 708) = 0.540	0.583	n.s.

Table 6.1.5 shows that there is no significant difference between the VOT values of labials and alveolars. Since affricates have significantly more aspiration than other stops, it

should be no surprise that they have significantly longer VOT than all other stops. Thus, it is only because of the inclusion of affricates and their treatment as a separate place of articulation that 6.1.5 shows place having a significant effect on VOT. Otherwise, the difference between labials and alveolars is not significant.

Table 6.1.5: Post-Hoc Analysis of Place of Articulation

Pairs	p =	Significance
Labial and Alveolar	0.438	n.s.
Labial and Affricate	< 0.01	*
Alveolar and Affricate	< 0.01	*

Going back to Table 6.1.4, the main factor of interest in this study, tone, turns out to have no significant effect on VOT. If aspirated and unaspirated stops are separated, however, we obtain different results. The effect of tone with unaspirated stops excluded was F = (1, 354) = 3.961, p = 0.047. Here, there is a significant difference between the VOT of the 55 and 33 tones for aspirated stops with the VOT of the 33 tone being higher.

6.2: Results for Part 2

Like in the results for Part 1 in the previous section, the results for Part 2 are also organized in a way to make it possible to look at the data in several different ways. Tables 6.2.1, 6.2.2, and 6.2.3 include all the average values of the VOT/Word Length ratio for Part 2.

Table 6.2.1: VOT/Word Length Averages for Aspirated Stops

Tone	Average	Std. Deviation	N
55	0.2487	0.06648	60
33	0.2619	0.06912	60
25	0.3002	0.06110	60
21	0.3240	0.07782	60
All Tones	0.2837	0.07476	240

Table 6.2.2: VOT/Word Length Averages for Unaspirated Stops

Tone	Average	Std. Deviation	N
55	0.0551	0.02382	60
33	0.0555	0.02834	60
25	0.0602	0.03353	60
21	0.0610	0.03357	60
All Tones	0.0580	0.03002	240

Table 6.2.3: VOT/Word Length Averages for Both Aspirated and Unaspirated Stops

Tone	Average	Std. Deviation	N
55	0.1519	0.10920	120
33	0.1587	0.11623	120
25	0.1802	0.13013	120
21	0.1925	0.14486	120
All Stops	0.1708	0.12651	480

ANOVA results for Part 2 are included in Table 6.2.4. This table shows that both aspiration and lexical tone have a significant effect on VOT. The fact that aspirated stops have significantly longer VOT is consistent with the results for Experiment 1 and once again it should be no surprise that we get the same results here. More importantly, however, is that Experiment 2 shows that there is a significant effect of tone on VOT. This is a result that was not obtained for Part 1 with the exception of the analysis of only the aspirated stops. A look at the Post-Hoc analysis in Table 6.2.5 shows exactly which tones are significantly different from each other.

Table 6.2.4: Significance of Factors on VOT/Word Length

Factor	F =	p =	Significance
Aspiration	(1, 472) = 2164	< 0.01	*; asp > unasp
Tone	(3, 472) = 15.029	< 0.01	*; 21, 25 > 33, 55
Aspiration*Tone	(3, 472) = 10.582	< 0.01	*

Table 6.2.5: Post-Hoc Analysis for All Tokens

Tonal Pairs	p =	Significance
55 & 33	0.32	n.s.
25 & 21	0.074	n.s.
55 & 25	< 0.01	*
55 & 21	< 0.01	*
33 & 25	< 0.01	*
33 & 21	< 0.01	*

According to 6.2.5 above, the results are still consistent with the results of Part 1 not counting the separate results for aspirated stops only. Both Part 1 and Part 2 show no significant difference between the 55 and 33 tones. Yet, with more tonal categories to compare in Part 2, we also have more data to help us understand how tone and VOT interact. Apparently, there is no significant difference either between the 25 and the 21 tones but all other possible pairs do show significant differences in VOT. If we were to organize all the tone categories on a scale that corresponds to the different VOT values associated with each tone, we would get the following:

(3) 21, 25 > 33, 55

Table 6.2.6 shows that even if we excluded the unaspirated stops as we did for Part 1, we get the same results showing tone to have a significant effect. We also get the same ranking of tones based on VOT values. One difference from the results of Part 1 without the unaspirated stops, however, is that Part 2 shows no significant difference between the 55 and 33 tones with or without unaspirated stops.

Table 6.2.6: Post-Hoc Analysis for Aspirated Stops Only

Tonal Pairs	p =	Significance
55 & 33	0.294	n.s.
25 & 21	0.061	n.s.
55 & 25	< 0.01	*
55 & 21	< 0.01	*
33 & 25	< 0.01	*
33 & 21	< 0.01	*

F = (3, 236) = 15.127, p < 0.01

7. Discussion

Based on the results of the experiment presented in this paper, we can conclude that tone does have a significant effect on VOT values. The results also make it clear that the tonal effects are patterned in a particular way. Since we saw no significant difference between the 55 and the 33 and between the 21 and 25 but we did see significant differences between all other pairings, we can see that there is a two-way division of the Cantonese tone categories. It appears that this two-way division is based on tonal onset as the 21 and 25 both have lower tonal onsets than do the 55 and 33 tones. It is interesting to note that this two-way division almost corresponds to the distinction between the High (*Yin*) and Low (*Yang*) tone registers presented in Table 4.1.2. The 25 tone, which is grouped with the High register, however, is grouped with a Low register tone, 21, based on VOT values. The overall pattern that emerges from the results is that <u>lower tones correspond to higher VOT.</u>

This pattern of lower tones associated with higher VOT seems to run counter to the generalization made by Maddieson (1978) that higher tones condition longer consonants. Results from this study show the opposite effects for bilabial stops. Then again, Maddieson also made this claim based on studies of fricatives and voiced stops. This current study looked at voiceless stops, so perhaps it does not necessarily contradict Maddieson's findings. It may be the case that different types of stops have different durational effects when they interact with different tones. In Part 1, we certainly did see a difference between the voiceless unaspirated and aspirated stops. It was only the aspirated stops that showed a significant difference with higher VOT for the lower tone. This is consistent with Kessinger and Blumstein's (1997) findings that voiceless unaspirated stops are generally stable in VOT values even at different rates of speech. Apparently, even across different tonal categories, VOT for unaspirated stops remains relatively stable. The fact that there is no significant difference between the 33 and 55 tones in Part 2 even if unaspirated stops are excluded is one finding that seems puzzling. A possible explanation could be the difficulty in measuring VOT for affricates. If this is the case, then that would mean that the significant difference between the 33 and 55 tones in Part 1 does not exist and that there were measurement errors made. If this is indeed the case, this would show that 33 and 55 really do group together and that they show no relationship of one tone corresponding to higher VOT values then for the other. Why they would group together would be the next question to ask.

One study that seems to support the results of this study is Lai (2004), which showed that higher VOT tends to be associated with lower F0 onsets. Comparing this project with Lai (2004) raises some interesting questions about the directionality of effects. Throughout much of this paper, she assumes that consonants affect the F0 onset rather than the F0 onset affecting the consonants. She does not explicitly say anything about the directionality of effects for the correlation of higher VOT with lower F0, though. So it seems possible that her study could be reinterpreted by considering the different possibilities of the direction of effects. The explanation she provides for why such a correlation exists seems to be a reasonable one. Longer VOT means that the perturbation effect of the aspiration decreases as time increases. Long aspirated stops have minimal effects on the following F0 onset. One question worth asking is why some aspirated stops would have longer VOT in the first place. The study presented in this paper suggests that perhaps it is because of the following tone that triggers the longer VOT when the tone is low. According to Gandour and Maddieson (1976), lower tone means a greater downward movement in the preceding consonant. With a greater downward movement in the larynx, this means that it takes longer for air to move out of the mouth. Consequently, aspirated stops produced with a lower larynx position means longer aspirated stops and hence longer VOT values.

Overall, this investigation of VOT and tone seems to support Maddieson’s claim that tones can have effects on preceding consonants. Then again, perhaps the real explanation to the observed phenomenon may be more complicated and may actually involve a number of coarticulatory factors. In other words, perhaps trying to frame the question in terms of whether the consonant is affecting the tone or whether the tone is affecting the consonant may be oversimplifying the situation. The reality may be that there are several factors involved at the same time that work together in intricate ways to yield the observed effects. Further research is needed before anything more conclusive can be said about why we see longer VOT associated with low tones.

8. Conclusion

To conclude, this study has shown that there is a correlation between lexical tone and VOT values. The overall pattern is that lower tones correspond to higher VOT. One possible direction for follow-up research would be to take F0 measurements of all of the data and see whether or not F0 onset values show as strong of a correlation with VOT values as tonal onsets do. F0 onset is not the same thing as tonal onset. While F0 onset is based on absolute values, which we do not have, tonal onset is something that we can say we assume as given. In other words, the tonal onset of a given token is a priori knowledge of what word the subject attempted to pronounce. It is quite clear from our data that there is a correlation of tonal onset to VOT values but we cannot conclude whether or not this is due to the tonal onset or to the F0 onset until we see whether or not F0 onset has as strong of an effect on VOT.

Acknowledgements

Acknowledgements go to my adviser Alan Yu for all of his patience, guidance, and constructive feedback; my second M.A. reader John Goldsmith for his support; Yuwen Lai, who helped tremendously with the statistical analysis; Fang Liu; Yi Xu for his feedback on my pilot study; my fellow classmates (Patrick Midtlyng, Jenn Watson, and Alison Wiener) in 'Experimental Phonetics', the course from which this project initially developed; and the six subjects who offered to participate in this study. 多謝!

References:

Boersma, Peter, and Weenink, David. 2005. *Praat 4.3.27: A system for doing phonetics by computer*. University of Amsterdam, Amsterdam, The Netherlands. Computer software available online: http://www.praat.org

Boucher, Victor. 2002. Timing Relations in Speech and the Identification of Voice Onset Times: A Stable Perceptual Boundary for Voicing Categories Across Speaking Rates. *Perception and Psychophysics* 64(1): 121-130.

Chao, Y.-R. 1930. A system of tone letters. *Le Maître Phonétique* 45: 24-27.

Chen, Matthew. 2000. *Tone Sandhi: Patterns Across Chinese Dialects*. Cambridge: Cambridge University Press.

Cho, Taehong. 2003. Lexical Stress, Phrasal Accent and Prosodic Boundaries in the Realization of Domain-Initial Stops in Dutch. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona.*

Cho, Taehong and Ladefoged, Peter. 1999. Variations and Universals in VOT: Evidence from 18 Languages. *Journal of Phonetics* 27, 207-229.

Ewan, W. G. 1976. Laryngeal Behavior in Speech. Doctoral Dissertation, University of California, Berkeley.

Gandour, Jack. 1974. Consonant Types and Tone in Siamese. *Journal of Phonetics* 2: 337-50.

Gandour, Jack and Ian Maddieson. 1976. Measuring larynx height in Standard Thai using the cricothyrometer. *Phonetica* 33: 241-267.

Halle, Morris and Kenneth Stevens. 1971. A Note on Laryngeal Features. *Quarterly Progress Report, Research Laboratory of Electronics*, MIT, 101: 198-213.

Hashimoto, Oi-kan Yue. 1972. *Studies in Yue Dialects 1: Phonology of Cantonese*. Cambridge University Press.

Haudricourt, André-Georges. 1954. De l'origine des tons en vietnamien. *Journal Asiatique* 24 : 69-82

Hombert, Jean-Marie. 1975. Towards a Theory of Tonogenesis: an Empirical, Physiologically and Perceptually-based Account of the Development of Tonal Contrasts in Language. Doctoral Dissertation, University of California, Berkeley.

Hombert, J. M. 1978. Consonant Types, Vowel Quality and Tone. In: V. Fromkin (ed.) *Tone: A Linguistic Survey*. 77-112. London: Academic Press.

Hombert, Jean-Marie and Peter Ladefoged. 1977. The Effect of Aspiration on the Fundamental Frequency of the Following Vowel. *UCLA Working Papers in Phonetics* 36: 33-40.

Hyman, Larry. 1976. On Some Controversial Questions in the Study of Consonant Types and Tone. *UCLA Working Papers in Phonetics* 33: 90-98.

Hyman, Larry and Russell Schuh. 1974. Universals of Tone Rules: Evidence from West Africa. *Linguistic Inquiry* 5: 81-115.

Keating, Patricia. 1984. Phonetic and Phonological Representation of Stop Consonant Voicing. *Language* 60: 286-319.

Kessinger, Rachel and Blumstein, Sheila. 1997. Effects of speaking rate on Voice Onset Time in Thai, French and English. *Journal of Phonetics* 25: 143-168.

Klatt, Dennis. 1975. Voice Onset Time, Frication, and Aspiration in Word-Initial Consonant Clusters. *Journal of Speech and Hearing Research* 18: 686-706.

Lai, Yuwen. 2004. The effect of aspiration on fundamental frequency in Taiwanese syllables. Unpublished M.A. Thesis. University of Kansas.

Lisker, Leigh and Arthur Abramson. 1964. A Cross-language Study of Voicing in Initial Stops: Acoustical Measurements. *Word* 20: 384-422.

Lisker, Leigh and Arthur Abramson. 1967. Some effects of context on voice onset time in English stops. *Language and Speech* 10: 1-28.

Maddieson, Ian. 1974. A Note on Tone and Consonants. *UCLA Working Papers in Phonetics* 27: 18-27.

Maddieson, Ian. 1976. A Further Note on Tone and Consonants. *UCLA Working Papers in Phonetics* 33: 131-159

Maddieson, Ian. 1978. Tone Effects on Consonants. *Journal of Phonetics*. 6: 327-344

Matthews, Steven and Virginia Yip. 1994. *Cantonese: A Comprehensive Grammar*. London: Routledge.

Repp, Bruno and Alvin Liberman. 1987. Phonetic Category Boundaries Are Flexible. In Harnad, Stevan (Ed.) *Categorical Perception: The Groundwork of Cognition*. p. 89-112. Cambridge: Cambridge University Press.

Summerfield, Quentin. 1981. Articulatory Rate and Perceptual Constancy in Phonetic Perception. *Journal of Experimental Psychology*, v. 7, no. 5, 1074-1095.

Van Dam, Mark. 2003. Voice Onset Time of American English Stops with Prosodic Correlates. *Poster Presentation at the Acoustical Society of America Conference*.

Xu, Yi and Ching X. Xu. 2003. Effects of Consonant Aspiration on Mandarin Tones. *Journal of the International Phonetic Association*, v. 33, 2: 165-181.

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

Zee, Eric. 1980. The Effect of Aspiration on the F0 of the Following Vowel in Cantonese. *UCLA Working Papers in Phonetics* 49: 90-97.

Zee, Eric. 1977. The effect of F0 on the duration of [s]. Paper presented to the *93rd Meeting of the Acoustical Society of America*, College Park, PA

Zee, Eric. 1999. "Change and variation in the syllable-initial and syllable-final consonants in Hong Kong Cantonese". *Journal of Chinese Linguistics*. v 27, 1: 120-167.