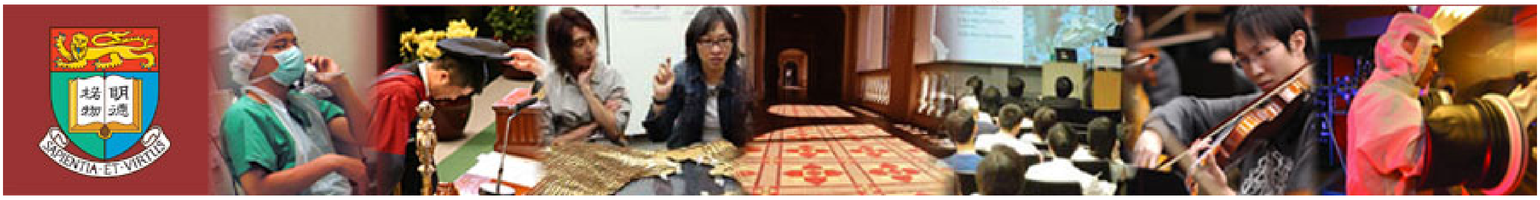

Title	Effect of tones on voice onset time (VOT) in Cantonese aspirated stops
Author(s)	Lam, Chung-ling; 林松齡
Citation	
Issued Date	2010
URL	http://hdl.handle.net/10722/173709
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Effect of tones on voice onset time (VOT) in Cantonese aspirated stops

Lam Chung Ling, Alfred

A dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2010

Effect of tones on voice onset time (VOT) in Cantonese aspirated stops

2006008236

A dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2010

Abstract

The study investigated the possible interaction between VOT values associated with aspirated stops produced at six different lexical tones (high falling, high rising, mid level, mid-low falling, mid-low rising and mid-low level) in Cantonese. A total of 27 male Cantonese speakers were recruited and they were instructed to read phrases containing targeted CV syllables formed by the aspirated Cantonese stops (/p^h/, /t^h/, and /k^h/) and the vowel /a/ at the six tones. VOT analysis revealed that, across aspirated stops, tones in the upper tone register produced shorter VOT while those in the lower tone register had longer VOT values. In particular, mid-low rising tone showed the longest VOT than all other tones. This finding indicated an interaction between VOT and tone during Cantonese stop production is confirmed.

Introduction

Voice onset time (VOT)

Voice onset time (VOT) of a plosive is defined as the time between the burst of air and the initiation of vowel sound during production of a plosive-vowel syllable (Lisker & Abramson, 1964). Physiologically, VOT corresponds to the duration between the release of an oral occlusion and the onset of vocal fold vibration. Based on the relative timing of voicing onset, VOT has been classified into four categories: voicing lead, simultaneous voicing, short lag and long lag (Kent & Read, 2002). When voicing starts before plosive consonant production, VOT is negative and it is called a voicing lead. For simultaneous voicing, voicing and release of stop occur at the same time. When voicing starts after the release of stop, VOT is positive, and it is labeled as either short lag or long lag depending on the VOT value (Kent & Read, 2002; Klatt, 1975).

As an acoustical measurement, VOT serves as an important perceptual cue for distinguishing different places of articulation, voicing and aspiration of plosive consonants (Cho & Ladefoged, 1999; Kent & Read, 2002). In a cross-linguistic study of VOT, Lisker and Abramson (1964) measured and compared VOT values of stops of different languages. In their investigation of Cantonese stops, they found that the three Cantonese aspirated stops (/p^h/, /t^h/, and /k^h/) were associated with longer VOT values (with a range of 77 - 87 ms) than unaspirated counterparts (/p/, /t/, and /k/) (with a range of 9 - 34 ms). VOT values of aspirated stops appeared to be long lags while those of unaspirated stops were short lags in standard Cantonese. In addition, the study also revealed the subtle relationship between VOT and place of articulation of stops. It was reported that when place of articulation changed from bilabial (/p/ & /p^h/) to velar (/k/ & /k^h/), the value of VOT gradually increased. For example, an aspirated bilabial stop /p^h/ had an average VOT of 9 ms, compared with 14 ms for the aspirated alveolar stop, and 34 ms for an aspirated velar stop /k/ (Lisker & Abramson, 1964).

Factors associated with VOT

Previous studies have suggested that several factors could influence VOT values of stops. Kessinger and Blumstein (1997, 1998) investigated the effect of speech rate on VOT and observed a significant change in VOT of voiceless stops with a change of speaking rate; an increase in speech rate could decrease VOT, at least for voiceless stops. Another factor that may affect VOT is speech task. Baran, Laufer, and Daniloff (1977) reported that, during conversational and reading tasks, the difference in VOT between voiceless and voiced stops reduced. In addition to speech rate and speech task, vowel context may also affect VOT values. Higgins, Netsell and Schulte (1998) found that longer VOT was associated with a stop consonant followed by the vowel /i/ than that by /a/. In another VOT study, Klatt (1975) reported that voiceless stop consonant had a longer VOT when followed by high vowels /i/ and /u/ than vowels /e/ and /ɛ/. The effect of vowel height on VOT may be related to the fact that both vocal fold tension and airflow could delay onset of vocal fold vibration and therefore lengthen VOT (Higgins et al., 1998).

Although effects of many factors on VOT have been well documented, effect of fundamental frequency (F0) on VOT is relatively lacking, especially in tonal language. F0 is an acoustical measurement corresponding to the rate of vocal fold vibration. In previous studies investigating the relationship between F0 and VOT in non-tonal language, McCrea and Morris (2005) conducted a study regarding the effect of F0 on VOT of English stops. They hypothesized a positive relationship between F0, phonation threshold pressure (PTP) and vocal fold tension (Löfqvist, Bear, McGarr, & Story, 1989; Titze, 1992). A stop consonant produced at high F0 was expected to result in a longer VOT because of the increased vocal fold tension and phonation threshold pressure (PTP) which delayed the onset of vocal fold vibration. To confirm their hypothesis, McCrea and Morris (2005) analyzed VOT values of six English voiced and voiceless stops (/b/, /d/, /g/, /p/, /t/ and /k/) produced by 56 young men using high, mid and low pitch levels. However, their results only partially supported what they hypothesized. The effect of F0 on VOT was only significant in voiceless stops, but not in voiced stops. In addition, across the three different F0 levels, the shortest VOT was noted for voiceless stops produced at high F0; intermediate VOT at mid F0; and the longest VOT at low F0 (McCrea & Morris, 2005). There appeared to be an inverse-proportional relationship between F0 and VOT values.

Seeing the lack of investigation of F0 effect on VOT in tonal language, Liu, Ng, Wan, Wang, and Zhang (2008) investigated the effect of tones on VOT between laryngeal and esophageal speakers of Mandarin. Liu et al. (2008) instructed eight laryngeal and seven esophageal speakers to produce syllables /p^ha/, /t^ha/, and /k^ha/ using high-level (HL), mid-rising (MR), falling-rising (FR), and high-falling (HF) tones. The results revealed significant differences in VOT values among the four different lexical tones produced by laryngeal speakers. HF tone was associated with shorter VOT values when compared with MR and FR tones. This indicates that changes in F0, especially an increase in pitch (as in rising tone), would lengthen VOT values. Liu et al. (2008) hypothesized that longer VOT found in MR and FR tones were related to voicing mechanism. They suggested that the rising component in MR and FR tones required greater vocal fold tension for a higher F0 and therefore, more time was needed for the anticipated increase in tension, which eventually delayed the onset of vocal fold vibration and thus yielded longer VOT values. This provided a new insight on the effect of F0 on VOT in tonal language.

Characteristics of Cantonese lexical tones

Similar to Mandarin, Cantonese is a tonal language in which words are contrasted by segmental information (consonants and vowels) and lexical tones (Bauer & Benedict, 1997). There are altogether nine lexical tones (T1 to T9) in Cantonese which are categorized into level (T1 & T4), elevating (T2 & T5), departing (T3 & T6) and entering (T7, T8 & T9). Each category is further divided into two register, upper and lower, which is determined by pitch value, F0 (Gu, Hirose, & Fujisaki, 2007). Taking an example (see Table 1), T2 has a relatively higher pitch value than T5 and therefore, T2 is categorized as high rising (HR), while T5 is categorized as low rising (LR). Some transcription schemes such as Jyut Ping (Deng, 2002) define only six Cantonese lexical tones by excluding the three entering tones (T7 to T9) as these entering tones show similar F0 with their non-entering counterparts (T1, T3 & T6).

Table 1

Description of Cantonese lexical tones

Traditional Cantonese tone names	Tone number	Pitch features
Entering tones		
Upper-level	T1	High level
Upper-elevating	T2	High rising
Upper-departing	T3	Mid level
Lower-level	T4	Low falling
Lower-elevating	T5	Low rising
Lower-departing	T6	Low level
Non-entering tones		
Upper-entering	T7	High level
Mid-entering	T8	Mid level
Lower-entering	T9	Low level

The six lexical tones in Cantonese, including high level (HL), high rising (HR), mid level (ML), low falling (LF), low rising (LR) and low level (LL), can be viewed as a linguistic description of the perceived pitch. Acoustically it is related to the fundamental frequency (F0), which corresponds to the rate of vocal fold vibration. Accordingly, Cantonese lexical tones are related to how rapidly the vocal folds vibrate and implicitly the vocal fold tension. To produce different tones, different vocal fold tensions are required and preparation time varies. As a result, tones should to some extent be related to the onset of vocal fold vibration and VOT values.

Aim of study

Although a few studies have investigated VOT characteristics of stops of tonal languages such as Mandarin, information regarding the interrelationship between VOT values and Cantonese lexical tones is still lacking. The present study aimed at examining the effect of Cantonese lexical tones on VOT values. Based on findings reported in the literature (Chen & Ng, 2005; Liu et al., 2008; McCrea & Morris, 2005), F0 and vocal fold tension are correlated with VOT. It is hypothesized that there are significant differences in VOT values of aspirated stops (/p^h/, /t^h/, and /k^h/) among the six Cantonese lexical tones. According to inverse-proportional relationship between F0 and VOT values (McCrea & Morris, 2005), upper tones (e.g., HL, HR, & ML tones) which have relatively high F0, would be expected to have shorter VOT values then lower tones (e.g., LF, LR, & LL tones). In addition, longer VOT values would be expected for HR and LR tones due to their rising nature of tone contours, as extra preparation time is needed for the increase in vocal fold tension which delays the onset of vocal fold vibration (Chen & Ng, 2005; Liu et al., 2008).

Method

Participants

A total of 27 male adults were recruited as participants for the study. They ranged in age from 19 - 25 years (mean age = 22 years). Only male participants were recruited in order to control for possible gender effect on VOT (Whiteside & Irving, 1997). All participants were physically healthy with no known history of speech or voice problems, laryngeal pathology, and/or hearing problem. They were all native speakers of Hong Kong Cantonese who were literate and able to read the speech materials used in the study.

Speech materials

To study the effect of lexical tone on VOT of stops, aspirated Cantonese stops (/p^h/, /t^h/, and /k^h/) followed by the vowel /a/ were used. The CV syllable was produced at six different tones of Cantonese: high falling (HF), high rising (HR), mid level (ML), mid-low falling (LF), mid-low rising (LR) and mid-low level (LL). The citation words consisted of real and non-existing words. The Chinese character of each of 18 target syllables (3 stops x 6 tones) was printed on cue cards, except the non-existing words which are lexically unavailable in Cantonese. The nonsense words were printed in IPA symbols as shown in Table 2. All citation words were embedded in a carrier phrase (printed in Chinese): "依個字係___" (meaning "This word is _______") in order to simulate a more natural production than isolated word production.

Table 2

Stimulus used in the present study (*: non-existing word)

Stops	Tones					
	HF	HR	ML	LF	LR	LL
/p^h/	趴	扒	怕	爬	*/$p^h a_5$/	*/$p^h a_6$/
/t^h/	他	*/$t^h a_2$/	*/$t^h a_3$/	*/$t^h a_4$/	*/$t^h a_5$/	*/$t^h a_6$/

/k^h/	卡	*/$k^h a_2$/	*/$k^h a_3$/	*/$k^h a_4$/	*/$k^h a_5$/	*/$k^h a_6$/

Recording procedure

The recording took place in a sound treated booth located at the Speech Science Laboratory, University of Hong Kong. Acoustic signals of the speech samples were recorded by using a high-quality microphone (SM58A, Shure) which was positioned approximately 10 cm in front of the mouth of participant via a preamplification unit (MobilePre USB, M-Audio). The recorded speech samples were digitized at 20 kHz and stored in a computer for later analyses.

Prior to the experiment, participants were provided with a brief practice period with the speech materials to familiarize themselves with the format of recording. This helped to ensure that the participants were able to give the best production during the actual recording. Each participant was instructed to produce the speech samples at a normal loudness and speech rate. During recording, the computer screen showing the whole phrase (carrier phrase + stimuli) was presented to participant and each phrase was repeated twice. Stimuli were presented randomly in order to avoid possible practice effects. For production of nonsense words, participants were asked to imitate investigator's production in order to ensure precise tone production. Upon completion of recording, a total of 972 speech samples (27 subjects x 18 targets x 2 trials) were recorded.

Data and statistical analyses

VOT values were obtained from the speech samples by using *Praat*, which is a signal analysis software. VOT value was determined as the time between the burst of air and the initiation of vowel sound during production of a plosive consonant. During analysis, the waveform and wideband spectrogram of each recorded speech sample were displayed on the screen. Two time markers were placed on the waveform to mark the release of air burst and the initiation of the following vowel which was indicated by the first identifiable period with the help of the second formant from the spectrogram. The temporal separation indicated by the two markers represented VOT values in milliseconds (See Figure 1).

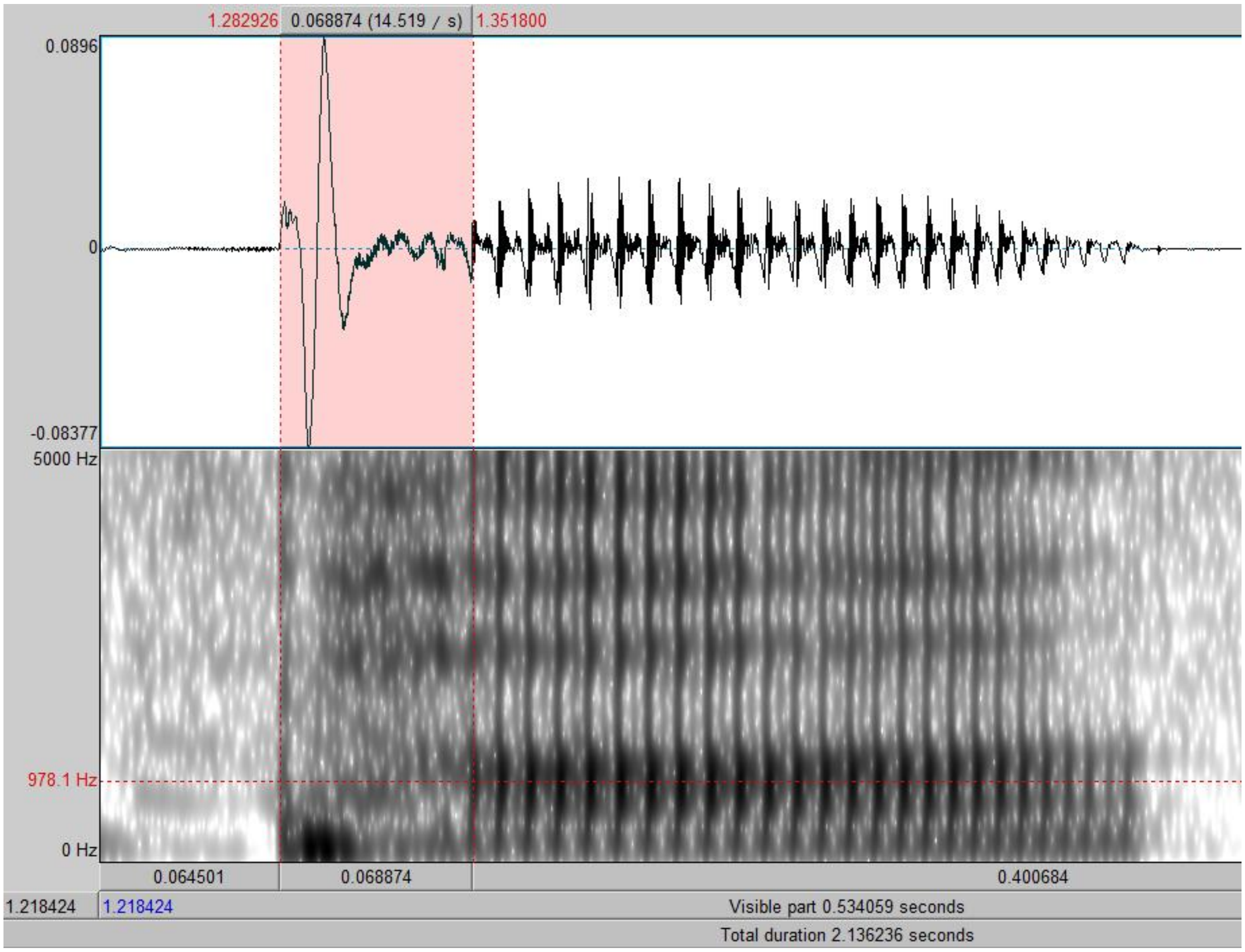

Figure 1. Spectrogram and waveform for Cantonese words “扒”.

As the study adopted a within-subjects design, repeated-measure Analysis of Variances (rm-ANOVA) was used to determine any significant differences present on VOT values across the six tones. Post hoc tests were conducted to determine specific VOT values difference across the six tones if needed.

Results

Reliability measurement

To evaluate intra-judge reliability, 5% (49 out of 972) of the recorded speech samples were randomly selected and analyzed by the same investigator again. To assess for inter-judge reliability, another set of 5% (49 out of 972) of the recorded speech samples was randomly selected and analyzed a second time by another investigator who was also experienced in measuring VOT values. The two re-measured sets of data were used to compare with the original one and absolute percent errors were calculated to determine intra-judge and inter-judge reliability.

The two sets of additional data were used to evaluate intra- and inter-judge reliability in comparison with the original data set. Absolute percent errors were used to indicate magnitude of the intra-judge and inter-judge reliability. Results showed an absolute error of 5.52% between the first and second measurements obtained by the primary investigator. An absolute error of 14.18% was obtained when comparing the data obtained by the primary investigator and secondary investigator. Both intra-judge and inter-judge reliability measurements reflected that primary investigator's analysis on VOT values was consistent and reliable.

VOT analysis

All recorded speech samples were analyzed by using *Praat*, and VOT values of all CV syllables are summarized in Table 3.

Table 3

Mean and standard deviation of VOT values (in ms) across the six Cantonese lexical tones and three aspirated stops produced by 27 subjects.

Stops	Tones					
	HL	HR	ML	LF	LR	LL

/p^h/	90.7 (13.05)	92.3 (13.59)	93.7 (11.79)	93.5 (15.29)	99.2 (20.31)	93.7 (17.14)
/t^h/	99.9 (14.27)	103 (14.28)	99.3 (12.68)	103 (10.56)	107 (14.86)	100 (14.10)
/k^h/	104 (16.63)	111 (14.72)	104 (13.94)	110 (14.52)	117 (16.79)	108 (10.65)

Repeated-measure Analyses of Variance (rm-ANOVAs) were used to determine the effect of Cantonese lexical tones on VOT values. Results of rm-ANOVAs revealed no significant interaction effect between the two independent variables: aspirated stops and lexical tones [$F(10,260) = 1.189$, $p = .298$]. However, significant main effect was found for tone [$F(5,130) = 6.661$, $p < .01$], which indicated that significant differences present on VOT values across the six Cantonese lexical tones. Moreover, significant main effect was also found for place of articulation of aspirated stops [$F(2,52) = 39.433$, $p < .01$]. This shows that VOT values across the three aspirated stops were significantly different. Aspirated velar stop /k^h/ had the longest VOT values (109.03 ± 14.54 ms) while aspirated bilabial stop /p^h/ had the shortest VOT values (93.86 ± 15.19 ms), with aspirated alveolar stop /t^h/ between the two (102.01 ± 13.46 ms) (See Figure 2).

Rm-ANOVAs also revealed that significant differences in VOT values were present across upper and lower tones [$F(1,26) = 11.549$, $p = .02$]. In Figure 3, it shows that longer VOT values (103.51 ± 12.66 ms) are found in lower tones (LF, LR and LL) than those of upper tones (HL, HR and ML) (99.76 ± 11.91 ms). Figure 4 shows the mean VOT values associated with the six Cantonese lexical tones obtained from the 27 speakers. It can be observed that low rising (LR) tone had the longest VOT values when compared with rest of the tones and this was consistent with the results from the Tukey's *post hoc* tests. In making pairwise comparisons, VOT values of LR tone were found to be significantly different from other tones, including HL tone (98.25 ±

14.65 ms) ($p < .02$), HR tone (102.18 ± 14.2 ms) ($p < .01$), ML tone (98.85 ± 12.8 ms) ($p < .02$), LF tone (102.15 ± 13.46 ms) ($p < .01$), and LL tone (100.7 ± 13.96 ms) ($p < .02$). In addition, another rising tone, HR tone, was found to be significantly different from ML tone ($p = .035$), but not in the other three tones: HL, LF and LL tones.

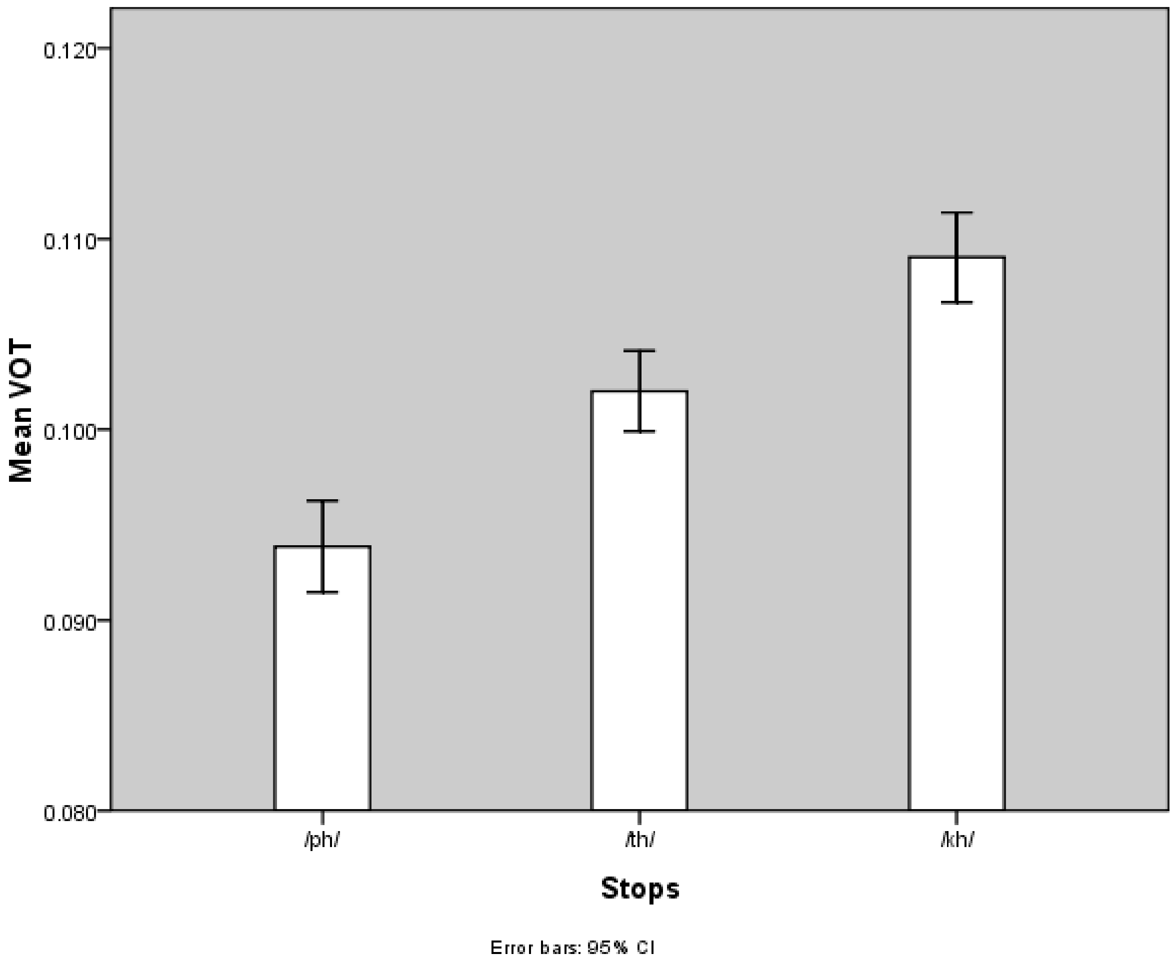

Figure 2. Average VOT values (in second) of three aspirated stops (/p^h/, /t^h/ and /k^h/).

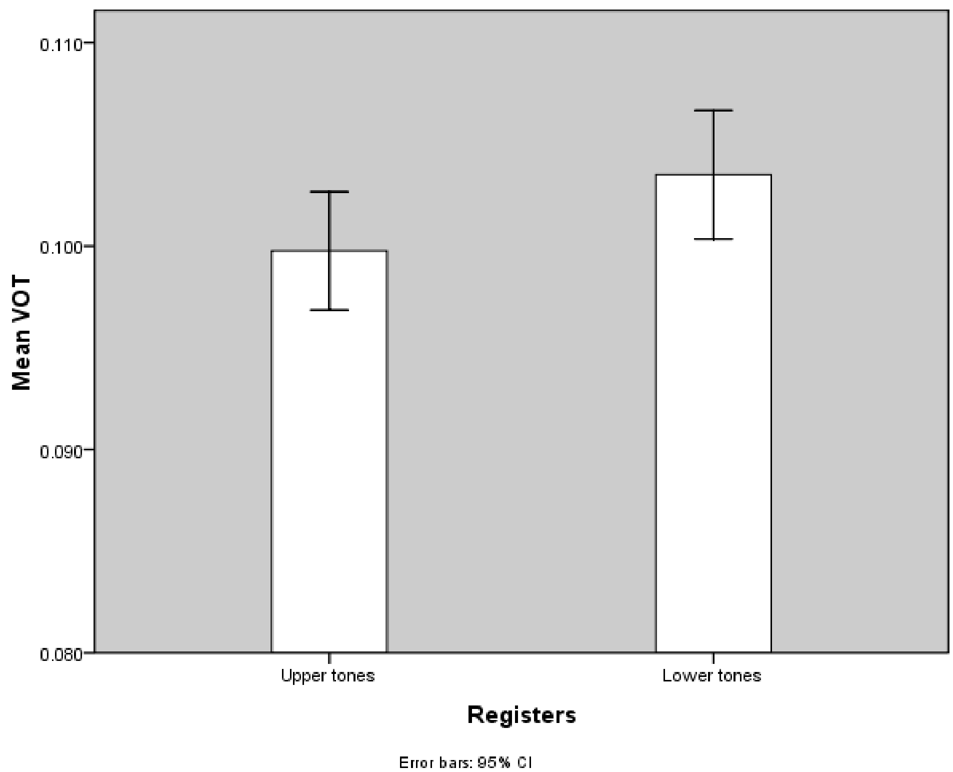

Figure 3. Mean VOT values (in second) associated with two registers: upper tones (HL, HR & ML) and lower tones (LF, LR & LL).

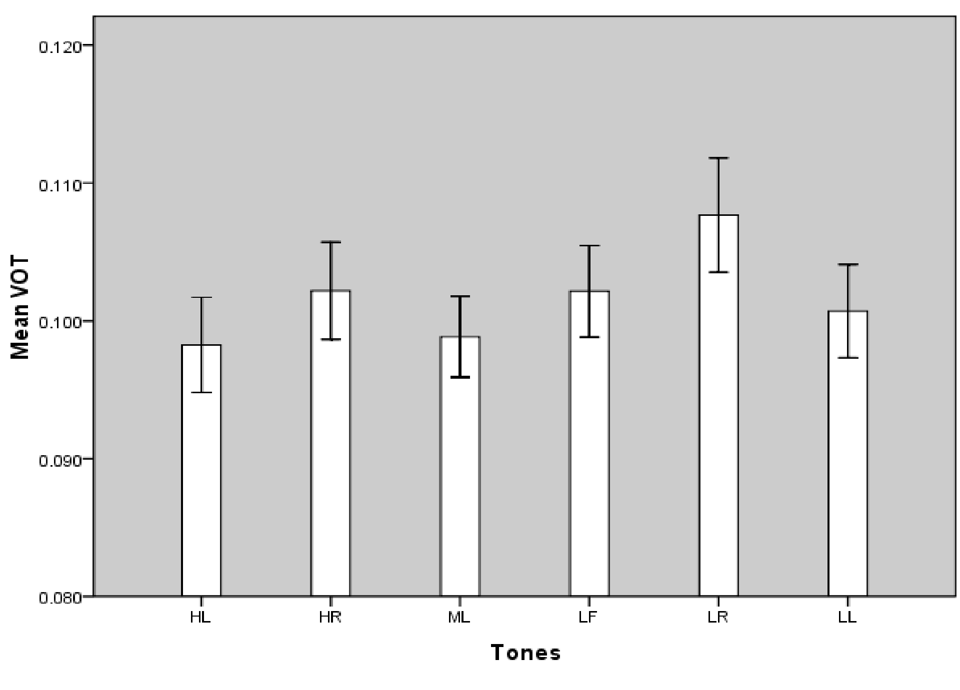

Figure 4. Mean VOT values (in second) associated with six lexical tones in 27 subjects

Discussion

Voice onset time (VOT) is an acoustical measurement which corresponds to the period between burst of air and the initiation of vowel sound during production of a plosive consonant (Lisker & Abramson, 1964). VOT also serves as a perceptual cue for distinguishing different places of articulation, voicing and aspiration of plosive consonants (Cho & Ladefoged, 1999, and Kent & Read, 2002). Moreover, other factors such as speech rate (Kessinger et al., 1997, 1998), speech task (Baran et al., 1977), and vowel context (Higgins et al., 1998) have been found to affect VOT. However, relatively little is know regarding the effects of fundamental frequency (F0) and tones on VOT. Whether and how F0 affects VOT is still unknown. The present study investigated the effect of tones on VOT of Cantonese aspirated stops. Results revealed three factors that may have an effect on VOT values: place of articulation, frequency (F0), and presence of a rising component in the pitch contour.

Place of articulation

The present results indicated that VOT value increased from bilabial to velar stops. A similar trend was reported by Lisker and Abramson (1964) in their cross-linguistic study. Common to many languages including Cantonese, English, Dutch, Spanish, Hungarian, Thai, Korean, etc., VOT of velar stops was the longest while that of bilabial stops was the shortest. Besides normal speaker, it was found that VOT value of stops produced by alaryngeal speakers also followed the trend. Christensen, Weinberg, and Alfonso (1978) investigated the VOT value of both voiced and voiceless stops produced by English esophageal speaker and concluded that VOT increased from bilabial to velar. Similar conclusion was also drawn from study on Mandarin (Liu, Ng, Wan, Wang, & Zhang, 2007), Thai (Gandour, Weinberg, Petty, & Dardarananda, 1987) and Cantonese alaryngeal speakers (Clumeck, Barton, Macken, &

Huntington, 1981).

When compared with bilabial and alveolar stops, velar stop takes relatively longer time to reduce intraoral air pressure to meet optimal pressure level for vocal fold vibration as oral occlusion of velar stop results a relatively small cavity which favors a more rapid increase of intraoral pressure (Maddieson, 1997). Secondly, Hardcastle (1973) reported the relationship between tongue dorsum mass, rate of jaw movement and onset of phonation. For velar stop, greater tongue dorsum mass and slower jaw opening movement are involved which delay the release of the oral occlusion in velar stop than alveolar and bilabial stop. This delays the time and rate of air burst after the release of velar stop occlusion and finally, velar stop requires longer time to meet the appropriate intraoral pressure for onset of phonation. Moreover, Stevens (1998) discovered that production of a velar stop is associated with a larger lingua-palatal contact area than alveolar and bilabial stops, and, with a wider oral constriction, articulators of velar stop will pull greatly together due to Bernoulli force. As a result, the increasing rate of the velocity of air flow through the constriction is reduced and then this lengthens the time required for dropping intraoral pressure to a condition that is favorable for onset of vibration. Overall, as more time is needed in velar stop for onset of vibration of vocal fold than alveolar and bilabial stop because of the longer time needed to achieve the required intraoral pressure after the release of stop occlusion. Therefore, the longest VOT value was observed in velar stop.

Frequency, F0

Löfqvist (1992) and Löfqvist, Baer, McGarr and Story (1989) hypothesized the relationship between fundamental frequency (F0), VOT value, vocal fold tension and phonation threshold pressure (PTP). Löfqvist et al. (1989) assumed that VOT value is directly proportional to F0. Production of stops in high F0 level requires increased vocal fold tension. When vocal fold tension increases with F0, more time is needed to meet the increased PTP required for initiation of vocal fold vibration when compared with stops produced with lower F0 level. However, the present result showed a different trend. HL, HR and ML tones in the upper tone register with higher F0 tended have shorter VOT values when compared with lower tones (LR, LF and LL tones). A similar trend has been observed in McCrea et al.(2005) in which VOT values associated with voiceless and voiced English stops across high, mid and low F0 level were determined. McCrea et al. (2005) concluded that VOT value is inversely proportional to F0 which disagreed with what Löfqvist et al. (1989) hypothesized.

The VOT pattern observed in the present study implies that VOT, F0, vocal fold tension and PTP do not interact as expected. Something must be missing which yielded the finding that shorter VOT at high F0 level. For successful vocal fold adduction, cricoarytenoid (CT) and thyroarytenoid (TA) play an important role (Hirano, Ohala, & Vennard, 1969, and Hiroto, Hirano, Toyozuma, & Shin, 1967). The muscles work together to tense the vocal folds and adduct by pulling them towards each other (Hirano et al., 1969). It is known that contraction of CT and TA muscles increases with F0 (Hirano et al., 1969; Shipp & McGlone, 1971), so greater contraction of muscles may result in faster vocal fold adduction rate and therefore, shorter time is needed for preparing stop production in high F0 level. McCrea et al. (2005) also suggested that stiffness of the vocal folds may be a factor contributing to shorter VOT at high F0 level. McCrea et al. (2005) hypothesized that greater vocal fold stiffness at high F0 is associated with narrowing of glottal width. With a reduced glottal width, the time needed for the vocal folds to completely adduct and start vocal fold vibration will be shortened, as the time required to build up sufficient subglottic pressure for phonation is reduced. Therefore, high stiffness of vocal folds and increased CT and TA contraction may be the contributing factors that cause shorter VOT at high F0 level.

Rising component of pitch contour

The present results indicate that rising tones including HR and LR tones have longer VOT than other tones within the same register. For example, in lower tone register, VOT in HR tone was longer than that in ML tone; in upper tone register, longer VOT was observed in LR tone when compared with those in LF and LL tones. A similar pattern was drawn from Liu et al. (2008), and Chen and Ng (2005) which investigated the effect of tones on VOT in Mandarin speakers. The present finding of the relationship between VOT and tone is consistent with the results reported previously. This implies that the rising component in a tone (such as HR and LR) appears to contribute to the longer VOT value observed. When producing rising tones, change in F0 along production is expected (Wang, Jongman, & Sereno, 2002). To prepare for such increase in F0, one needs to tense the vocal folds gradually, yielding a delayed onset of vocal fold vibration, and thus a longer VOT value. This is in line with Finkelhor, Titze, and Durham (1987), in which increased vocal folds tension was found to be associated with longer VOT value. Liu et al. (2008) further explained that this anticipated changes in vocal fold tension and F0 level of rising tones lengthen the preparation time for initiation of phonation inside the vocal tract. Furthermore, data from Xu and Sun (2002) showed that longer VOT is observed when participants increase the F0 rather than decrease it. This further supports the hypothesis that pitch rising needs greater voicing loading than pitch falling. In order to meet this requirement of greater voicing loading, more time is needed to prepare our vocal tract before increasing F0 together with tension along vocal folds during rising tone production.

Limitations of the study

Due to time limitation, only 27 male subjects were recruited. Without involving a comparable number of female subjects, possible gender effect on VOT could not be drawn

(Whiteside & Irving, 1997). Moreover, a study involving both male and female speakers should yield results that are more representative and generalizable. In determining the effect of tones on VOT, acoustic analysis was carried out in the study. However, acoustic analysis only would solely give us the insight on the relationship between tones and VOT, but not the underlying mechanism of the phenomenon observed, at least not directly. Therefore, other analytical instrumental analyses such as laryngeal imaging are needed for providing information regarding structural changes, e.g., glottal width, vocal fold length, during stop production across different tones. In addition, aerodynamic and electromyographic studies would also be beneficial in determining the underlying mechanism which allows longer VOT in rising tones by providing supporting information on air flow, glottal pressure and muscles contraction.

Furthermore, in order to resemble a more natural production than isolated word production, all targeted words were embedded in a carrier phrase in the current study. However, naturalness of speech production can still be improved by involving target syllables in various vowel contexts, for facilitating generalization to spontaneous speech. Also, with target syllables in various vowel contexts, we can have a better understanding on how VOT interact with vowels and tones.

Conclusion

Results of the present study indicate that tone is another contributing factor that can affect VOT. It was observed that tones in upper tone register have shorter VOT than tones in lower tone register. Tones with a rising component such as HR and LR tones were observed to exhibit a longer VOT than other tones. It is believed that stiffness of vocal fold, glottal width, and greater demand in voicing mechanism contribute to the different VOT values across the six lexical tones in Cantonese. Further studies should investigate VOT across genders to check whether present results are consistent in both male and female. Also, in order to have a clearer picture on the underlying reason of tonal effect on VOT, aerodynamic, electromyographic or laryngeal imaging study should be carried out.

Acknowledgments

I would like to give special thanks to Dr. Manwa L. Ng for his support and help provided during the whole research process. I am also thankful to all participants for their help in the data collection process and contribution to the whole study.

References

Baran, J. A., Laufer, M. Z., & Daniloff, R. (1977). Phonological Contrastivity in conversation: A comparative study of voice onset time. *Journal of Phonetic, 5*, 339-350.

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. New York: Mouton de Gruyter.

Cho, T & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics, 27*, 207-229.

Chen, Y., & Ng, L. M. W. (2005). Examination of voicing onset time during Mandarin tone productions. *J Acoust Soc Am., 117*,2459.

Christensen, J. M., & Weinberg, B., & Alfonso, P. J. (1978). Productive voice onset time characteristics of esophageal speech. *Journal of Speech and Hearing Research, 21*, 56-62.

Clumeck, H., Barton, D., Macken, M. A, & Huntington, D. A. (1981). The aspiration contrast in Cantonese word-initial stops: Data from children and adults. *Journal of Chinese Linguistics, 9*, 210-224

Gandour, J., Weinberg, B., Petty, S. H., Dardarananda, R. (1987). Voice onset time in Thai alaryngeal speech. *Journal of Speech and Hearing Disorders, 52*, 288-294.

Gu, W., Hirose, K., & Fujisaki, H (2007). Analysis of tones in Cantonese speech based on the command-response model. *Phonetica, 64*, 29-62.

Deng S. (2002). *Yue yu pin yin zi biao (粵語拼音字表)* (2nd ed.). Hong Kong: Linguistic Society of Hong Kong.

Finkelhor B. K., Titze I. R., Durham P. I. (1987). The effect of viscosity changes in the vocal folds on range of oscillation. *J Voice, 1*, 320–325.

Hardcastle, W. J. (1973). Some observations on the tense-lax distinction in initial stops in Korean.

Journal of Phonetics, 1, 263-271.

Higgins, M. B., Netsell, R., & Schulte, L. (1998). Vowel related differences in laryngeal articulatory and phonatory function. *Journal of Speech, Language, and Hearing Research, 41,* 616-628.

Hirano, M., Ohala, J., & Vennard, W. (1969). The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation. *Journal of Speech and Hearing Research, 41*, 616-628.

Hiroto, I., Hirano, M., Toyozuma, Y., & Shin, T. (1967). Electromyographic investigation of the intrinsic laryngeal muscles related to speech sounds. *Annals of Otology, Rhinology, and Laryngology, 76,* 861-872.

Kessinger, R., & Blumstein, S. (1997). Effects of speaking rate on voice onset time in Thai, French, and English. *Journal of Phonetics, 25*, 143-168.

Kessinger, R., & Blumstein, S. (1998). Effects of speaking rate on voice onset time and vowel production: Some implications for perception studies. *Journal of Phonetics, 26*, 117-128.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonants clusters. *Journal of Speech and Hearing Research, 18*, 686-706.

Kent, R. D., & Read, C. (2002). *Acoustic analysis of Speech* (2nd ed.). Australia: Thomson Learning.

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

Liu, H. J., Ng, L. M. W., Wan, M. X., Wang, S. P., & Zhang, Y. (2007). Effects of place of articulation and aspiration on voice onset time in Mandarin esophageal speech. *Folia Phoniatrica et Logopaedica, 59*(3), 147-154.

Liu, H. J., Ng, L. M. W., Wan, M. X., Wang, S. P., & Zhang, Y. (2008). The effect of tonal changes on voice onset time in Mandarin esophageal speech. *Journal of Voice*, *22*, 210-218

Löfqvist, A. (1992). Aerodynamic measurements of vocal function. In A. Blitzer, M. F. Mitchell, C. T. Sasadi, S. Fahn, & K. S. Harris (Eds.). Neurologic disorders of the larynx. New York: Thieme Medical.

Löfqvist, A., Baer, T., McGarr, N. S., & Story, R. S. (1989). The cricothyroid muscle in voicing control. *Journal of Acoustical Society of America*, *85*, 1314-1321

Maddieson, I. (1997). Phonetic universals. In J. Laver & W. J. Hardcastle (Eds.). *The handbook of phonetic sciences*. United Kingdom: Blackwell.

McCrea, C. R., & Morris, R. J. (2005). The effects of fundamental frequency level on voice onset time in normal adult male speakers. *Journal of Speech, Language, and Hearing Research, 48*, 1013-1024.

Shipp, T., & McGlone, R. E. (1971). Laryngeal dynamics associated with voice frequency change. *Journal of Speech and Hearing Research, 14*, 761-768.

Stevens, K. N. (1998). Acoustic phonetics. Cambridge, MA: MIT Press.

Titze, I. R. (1992). Phonation threshold pressure: A missing link in glottal aerodynamics. *Journal of the Acoustical Society of America, 91*, 2926-2935.

Wang, Y., Jongman, A., Sereno, J, A. (2002). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of Americ*,113, 1033-1043.

Whiteside, S. P., & Irving, C. J. (1997). Speakers' sex differences in voice onset time: Some preliminary findings. *Perceptual and Motor Skills, 85*, 495-463.