聲調對嗓音起始時間的影響：以國語和客語為研究對象
Tonal effects on voice onset time: Stops in Mandarin and Hakka

彭瑞鳳 Jui-Feng Peng
陳麗美 Li-mei Chen
林依雲 Yi-Yun Lin

國立成功大學外國語文學系（所）
Department of Foreign Languages & Literature
National Cheng Kung University
leemay@mail.ncku.edu.tw

Abstract

This study examines the influence of lexical tone upon voice onset time (VOT) in Mandarin and Hakka. Examination of VOT values for Mandarin and Hakka word-initial stops /p, t, k, p^h, t^h, k^h/ followed by three vowels /i, u, a/ in different lexical tones revealed that lexical tone has a significant influence on the VOTs. The result is important because it suggests that future studies should take its influence into account when studying VOT values for stops in tonal languages. In Mandarin, stops' VOTs, ordering from the longest to the shortest, are in Tone 2, Tone 3, Tone 1, and Tone 4: this sequence is the same as Liu, Ng, Wan, Wang, and Zhang's (2008) [1] results. However, later it was found that the sequence results from the existence of non-words. Because in order to produce non-words correctly, participants tended to pronounce them at a lower speed, especially those in Tone 2. Therefore, we further examined the data without non-words, in which no clear sequence had been found. For Hakka, Post hoc tests (Scheffe) show that aspirated stops in Tones 4 and 8 have significantly shorter VOT values than they have in other tones.

Keywords: Voice onset time, Mandarin tones, Hakka stops, Mandarin stops

1. Introduction

The aim of this paper is to explore whether lexical tones influence the VOT values for word-initial stops. This issue is important because VOT is considered as a reliable phonetic feature to differentiate consonant stops ([2], [3], [4], [5], [6], [7]) and recently it has been used to study the language production of patients with language deficits or disorders ([8], [9]). Among the languages being investigated, some are tone languages, i.e. Mandarin, Cantonese, and Taiwanese. In a tonal language, the duration of each lexical tone is slightly different. Consequently, it is possible that lexical tone will affect stop's voice onset time. However, few studies have taken this factor into consideration while studying tone languages. It is hoped that with data from Mandarin and Hakka, we can establish the groundwork for future studies related to VOTs in tonal languages. If lexical tone does have an influence on the VOT, it should be taken into account when creating stimulus words in future studies for tonal languages, thereby rendering studies more valid and reliable.

1.1 Voice onset time

Lisker and Abramson (1964) [2] have defined voice onset time (VOT) as the temporal interval from the release of an initial stop to the onset of glottal pulsing for a following vowel. It has been considered as a reliable phonetic cue to categorizing the stop consonants, i.e. voiced vs. voiceless or unaspirated vs. aspirated, in various languages ([2], [3], [4], [5], [6], [7], [10]). Additionally, by comparing VOT values for stops produced by native and non-native speakers for specific languages, researchers have provided some suggestions for language learning and teaching ([6], [11], [12]). Moreover, recently researchers have studied aphasia, apraxia and stuttering patients' production deficits by observing their VOT values for stops ([8], [9]).

1.2 Factors affecting voice onset time

When investigating stops, researchers found that the VOT values for stops varied in relation to the place of articulation. Cho and Ladefoged (1999) [4], sorted out researchers' findings, have claimed that the further back the closure, the longer the VOT ([2], [4], [6], [13]). That is velar stops have the longest VOT values, alveolar stops the intermediate values, and bilabial stops have the shortest values. However, there are some exceptions. Alveolar stops in Tamil, Cantonese, Eastern Armenian, Hungarian, Japanese, and Mandarin, have shorter VOTs than bilabial stops ([2], [3], [5], [7], [12], [14]).

Liu et al. (2008) [1] speculated that the VOT durations may be affected by tone, because different tones have different fundamental frequencies and pitch levels, which are determined mainly by the tension of the vibrating structure. In order to achieve different levels of tension, different amounts of time might be needed. Consequently, the VOT values may vary when they are in different lexical tone. Only a few studies have tried to examine whether lexical tone influences VOT values. For example, Liu et al. (2008) [1] studied the effect of tonal changes on VOTs between normal laryngeal and superior esophageal speakers of Mandarin Chinese, and reported that for normal laryngeal speakers there are significant differences of VOT values caused by lexical tones. In addition, stops in Tone 4 have significantly shorter mean VOT values than stops in Tones 2 and 3. The study by Liu et al. [1] is a pioneering piece of work in this field, but more evidence is still needed. Therefore, by carrying out a systematic study with respect to the influence of lexical tone for stop's VOT using two tonal languages, i.e. Mandarin and Hakka, we try to verify previous findings in order to provide references for future linguistic studies on tonal languages.

1.3 The features of Mandarin and Hakka

Mandarin Chinese and Hakka are tonal languages, in which a word's meaning can be changed by the tone in which it is pronounced. Chao (1967) [15] suggested a numerical notation for lexical tones: dividing a speaker's pitch range into four equal intervals by five points: 1 low, 2 half-low, 3 middle, 4 half-high, and 5 high. The numerical notation indicates how the pitches of a lexical tone change. For example, the numerical notation for Tone 2 in Mandarin is 35, which represents that the pitch will go from middle to high. Table 1 reveals the numerical notation for each lexical tone in Mandarin and Hakka. In Mandarin, there are four contrasting lexical tones, Tone 1 (high–level), Tone 2 (mid-rising), Tone 3 (falling-rising), and Tone 4 (high-falling). Sixian Hakka has six contrasted lexical tones, Tone 1 (24), Tone 2 (31), Tone3 (55), Tone 4 (<u>32</u>), Tone 5 (11), and Tone 8 (<u>55</u>). The pitch values for Tone 3 and Tone 7 are the same, therefore Tone 7 has been omitted. Although there are regional differences for Hakka, Sixian Hakka was chosen as it is the most widely used Hakka dialect in Taiwan.

Table 1. The numerical notations for lexical tones in Mandarin [15] and Hakka [16].

Lexical Tone		1	2	3	4	5	7	8
Numerical notation	Mandarin	55	35	214	51			
	Hakka	24	31	55	32	11	(55)	55

Note: Those which are underlined represent pitches that are short and rapid.

Mandarin Chinese and Hakka have their specific tone sandhi rules and one example from each language is listed below. In Mandarin, Tone 3, which has the longest duration among the four lexical tones, will become Tone 2 while it is followed by another Tone 3 [17]. The tone sandhi rule for Sixian Hakka is as follows: Tone 1 will become Tone 5, when it precedes Tone 1, Tone 3, or Tone 8 [18]. Therefore, tone sandhi rules are taken into consideration when making stimulus words, and the combinations that might cause tonal change will be avoided.

Mandarin
Tone 3 →Tone 2 / _____ { Tone 3 }

Sixian Hakka
Tone 1→Tone 5 / _____ { Tone 1, Tone 3, Tone 8 }

2. Methodology

Mandarin and Hakka word-initial stops, unaspirated /p, t, k/ and aspirated /p^h, t^h, k^h/, in combination with three vowels /i, u, a/ were studied. Except for participants and stimulus words, the methodology employed for both languages was the same.

2.1 Participants

Mandarin and Hakka participants were different. For Mandarin, there were fifteen male and fifteen female native speakers recruited from college students and staff from an elementary school in Tainan City. All the participants grew up in Taiwan, with no hearing and speech defects. Their ages ranged from 23 to 33 years (mean = 27.2 years). As for the Hakka, there were twenty-one participants, eleven men and ten women, from Miaoli, Pingtung, and Taoyuan County. Their average age was fifty-one, the oldest being eighty, and the youngest thirty-six. As it was not easy to find fluent Hakka speakers their age range was quite wide.

2.2 Stimuli and procedure

The speech stimuli in both language were combination of six stops /p, t, k, p^h, t^h, k^h/ and three vowels /i, u, a/, i.e., 18 combinations. They were /pi/, /pu/, /pa/, /ti/, /tu/, /ta/, /ki/, /ku/, /ka/, /p^hi/, /p^hu/, /p^ha/, /t^hi/, /t^hu/, /t^ha/, /k^hi/, /k^hu/, and /k^ha/. For Mandarin there were four contrasting lexical tones, thus 72 monosyllabic words were created in total. Among them, 18 combinations do not exist in Mandarin. As for Hakka, there were six contrasting lexical tones in Sixian Hakka, hence there were 108 monosyllabic words obtained. Among these stimulus words, 12 words do not actually exist in Hakka. Chen et al. (2007) [14] has claimed that disyllabic words can create a more natural-like context for participants. Therefore, in order to make speakers produce the words more naturally, all the stimulus words were followed by another word and would become meaningful disyllables. For example, Mandarin word, /pi/, was followed by another word, /p^huo/ to become the existing disyllable, /pi p^huo/

(force). Some stimulus words in Hakka were tri-syllabic, due to the fact that no meaningful disyllables were found.

The stimulus words were arranged randomly, and the participants were asked to read it out loud at a normal speed. After finishing, the participants were asked to read out the words for the second time. Therefore, two groups of data were gathered for each participant. All the speech was recorded by a 24 bit WAV recorder, connected with a AKG head-worn cardioid condenser vocal microphone positioned of approximately 10~15 cm from the participant's mouth in a quiet room.

2.3 Data Measurement and analysis

After recording, data were edited into individual files and analyzed using the Praat software. VOT, measured in milliseconds (ms), was obtained by measuring the temporal interval between the beginning of the release burst and the onset of the following vowel as shown in Figure 1. The values of both the waveform and spectrogram were recorded, but the VOTs were determined primarily through waveform analysis; the values in the spectrogram were provided as references. If the values in waveform differed from the values in the spectrogram by more than five milliseconds, the data were re-measured to verify accuracy.

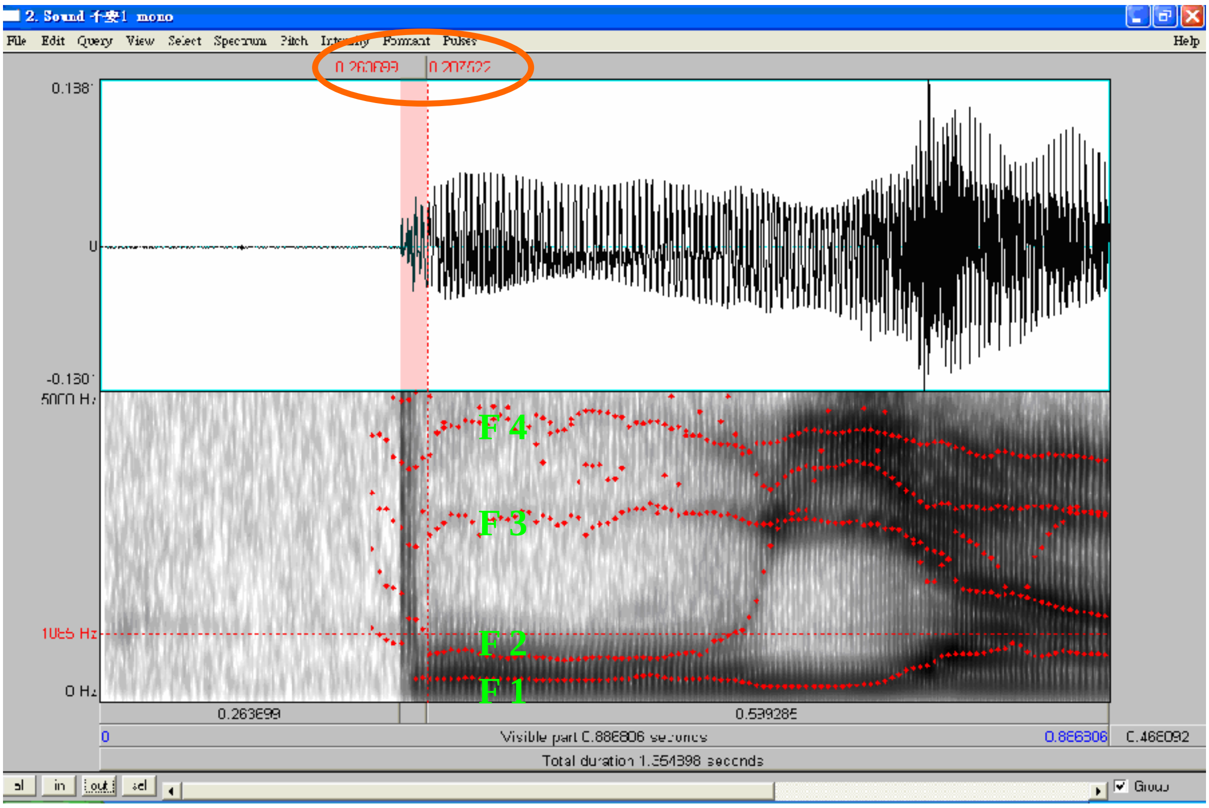

Figure 1. The spectrogram and waveform for the Mandarin word /pu iau/ 'don't want'. The values in the circle are the starting and endpoints of the VOT in the spectrogram.

When analyzing the data, the VOT values for the mispronounced words were excluded, and the data for Hakka /pi/ in Tone 8 were not analyzed because of wrong word-choosing. ANOVA test was used to examine whether or not there is a significant influence on stop's voice onset time. In addition, the differences between the examined targets were analyzed by Post Hoc tests (Scheffe). The measurements of stops' VOT values were made by the same investigator. Furthermore, randomly selected 10% of each recording were re-measured by another investigator to verify the reliability of the results. Therefore, 7 Mandarin words and 11 Hakka words for each recording were re-measured. The inter-rater reliability was then examined by Pearson's product-moment correlations.

3. Results

Pearson's product-moment correlations indicated high inter-rater agreement for both the Mandarin and Hakka data (Mandarin: r = .995, p<.001; Hakka: r = .978, p<.001). This indicates that the measurements were reliable throughout. It was found that the mean VOTs for Mandarin stops get longer due to the existence of non-words. Therefore, the data excluding non-words was further examined to verify the results. For Hakka, there is no clear difference because most of the non-words were pronounced incorrectly. Therefore, most of the values of Hakka non-words are not included in the analysis.

3.1 Lexical tone and VOT in Mandarin

Mandarin stops' mean VOT values and standard deviations in each lexical tone are shown in Table 2. ANOVA test reveals that lexical tones have significant influences on the VOTs for stops ($F(3,1040)=2.681$, p<.05 for unaspirated stops; $F(3,1040)=8.934$, p <.001 for aspirated stops). When examining the data with non-words, it is shown that for both unaspirated and aspirated, stops in Tone 2 have the longest mean VOTs and stops in Tone 4 have the shortest mean VOTs. Stops' VOT values ordering from the longest to the shortest are in Tone 2, Tone 3, Tone 1, and Tone4. Post hoc tests revealed that aspirated stops in Tone 4 have significantly shorter mean VOTs than stops in Tone 2 and Tone 3 (p< .05).

Table 2. Mandarin stops' mean VOT values in individual lexical tones. All measurements are in milliseconds (ms).

	With non-words				Without non-words			
	unaspirated stops		aspirated stops		unaspirated stops		aspirated stops	
	mean	*SD*	mean	*SD*	mean	*SD*	mean	*SD*
Tone 1	20.20	(11.90)	92.72	(25.53)	17.71	(9.95)	88.69	(20.4)
Tone 2	21.10	(12.68)	101.02	(30.21)	13.99	(6.03)	89.47	(23.31)
Tone 3	20.89	(13.35)	97.03	(27.75)	17.00	(10.98)	92.30	(23.49)
Tone 4	18.42	(9.94)	89.4	(25.72)	16.32	(9.07)	85.62	(24.18)

The results were verified by examining the data without non-words. In Figures 2 and 3, it is noted that the values for the data without non-words are shorter than the values for the data with non-words. It additionally shows that unaspirated stops in Tone 1 have the longest mean VOT values and unaspirated stops in Tone 2 have the shortest. Aspirated stops in Tone 3 have the longest mean VOTs, while those in Tone 4 have the shortest. ANOVA tests still indicate that lexical tone has significant influences on the VOT values for stops ($F(3,692)=4.800$, p<.01 for unaspirated stops; $F(3,779)=2.953$, p<.05 for aspirated stops). Furthermore, Post Hoc tests show that unaspirated stops in Tone 2 have significantly shorter mean VOTs than stops in Tone

1 and Tone 3 (p<.05), and aspirated stops in Tone 4 have significantly shorter mean VOTs than stops in Tone 3 (p<.05).

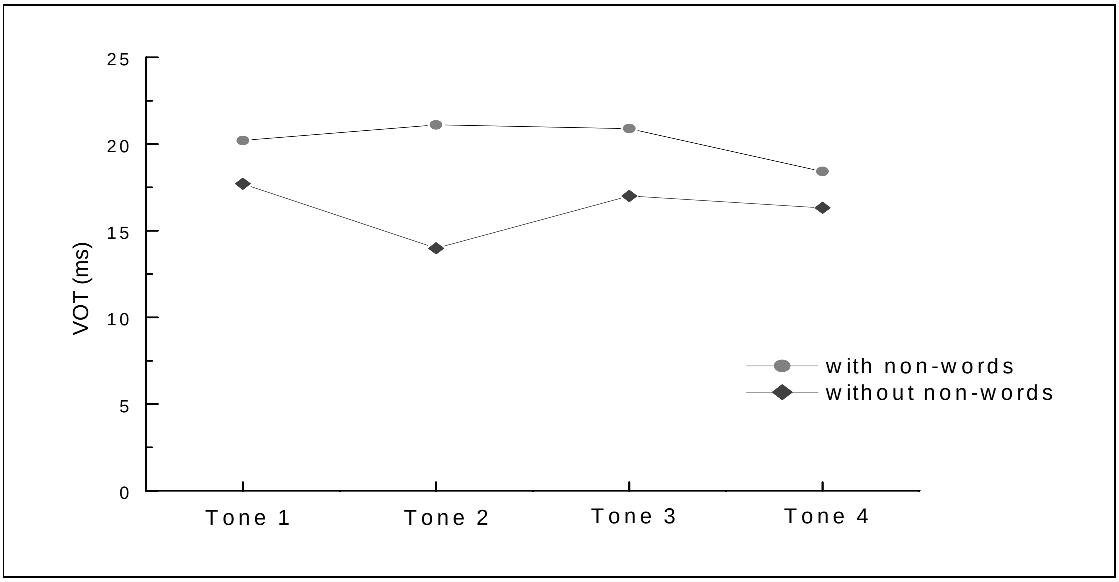

Figure 2. The mean VOTs for Mandarin unaspirated stops in individual lexical tones.

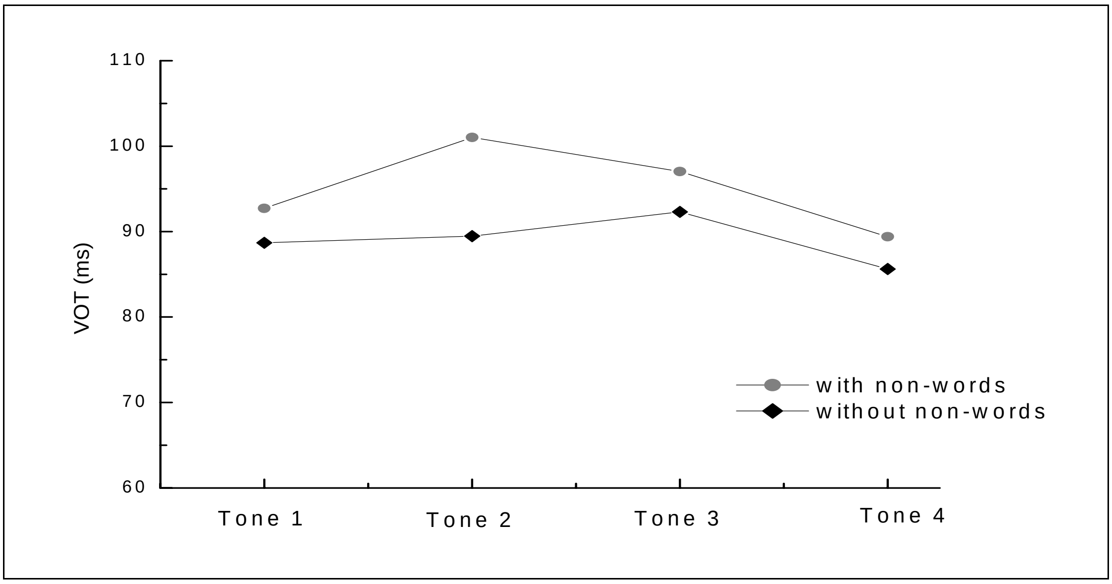

Figure 3. The mean VOTs for Mandarin aspirated stops in individual lexical tones.

3.2 Lexical tone and VOT in Hakka

The mean VOT values and standard deviations for Hakka stops in each lexical tone are shown in Table 3. ANOVA tests show that lexical tones have a significant influence on stop's VOTs ($F(5,943)=3.521$, $p<.01$ for unaspirated stops; $F(5,900)=37.365$, $p<.001$ for aspirated stops). In Figures 4 and 5, it is shown that unaspirated and aspirated stops in Tone 1 and Tone 5 have longer mean VOTs than stops in other tones. And the shortest mean VOTs for both unaspirated and aspirated stops are in Tone 8. Post hoc tests revealed that aspirated stops in Tone 4 and Tone 8 have significantly shorter mean VOTs than in Tone 1, Tone 2, Tone 3, and Tone 5 (p<.001).

Table 3. Hakka stops' mean VOT values in individual lexical tones. All measurements are in milliseconds (ms).

	Unaspirated stops		Aspirated stops	
	mean	(*SD*)	mean	(*SD*)
Tone 1	20	(11.56)	86.83	(25.8)
Tone 2	16.94	(8)	84.67	(26.56)
Tone 3	18.88	(11.02)	81.32	(23.73)
Tone 4	17.19	(9.44)	62.93	(18.36)
Tone 5	19.4	(11.43)	90.08	(27.08)
Tone 8	16.11	(7.98)	61.53	(20.36)

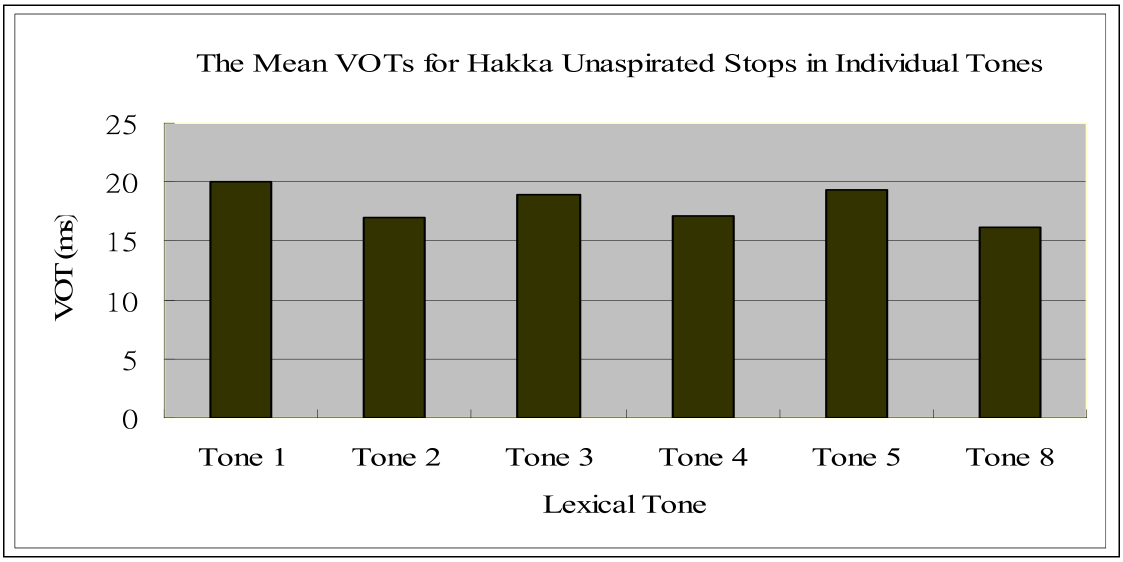

Figure 4. The mean VOTs for Hakka unaspirated stops in individual lexical tones.

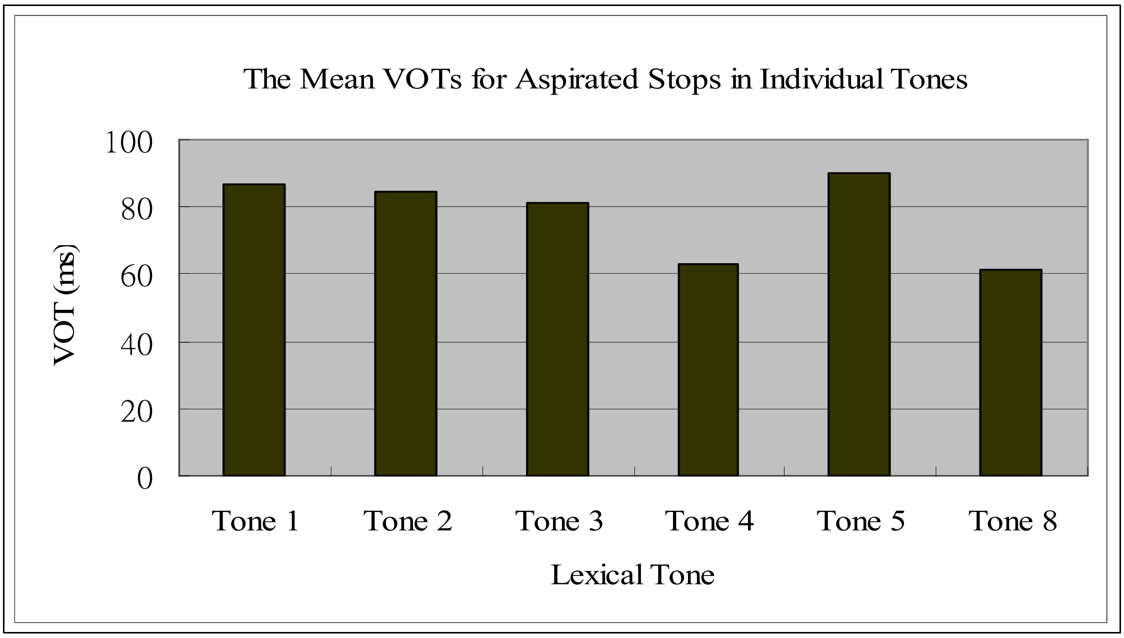

Figure 5. The mean VOTs for Hakka aspirated stops in individual lexical tones.

4. Discussion and Conclusion

In the current study, ANOVA tests reveal that lexical tone has a significant influence on the VOT values for Mandarin and Hakka stops. In Mandarin, the VOTs for both unaspirated and aspirated stops in Tone 2 have the longest mean VOT values, and in Tone 4 have the shortest mean VOT values. Stops' VOTs, ordering from the longest to the shortest, are in Tone 2, Tone 3, Tone 1, and Tone 4. This sequence is the same as Liu et al.'s (2008) [1] results. But it is worth noting that in both studies, some of the stimulus words are non-words. Later, it was found that the sequence results from the existence of non-words because in order to produce non-words correctly, participants tended to pronounce them at a lower speed, especially those in Tone 2. Therefore, we examined the data without non-words, in which no clear sequence had been found. In general, ANOVA tests revealed that lexical tones have significant influences on stops' VOTs. Moreover, Post hoc tests show that unaspirated stops in Tone 2 have significantly shorter mean VOTs than in Tones 1 and 3; while aspirated stops in Tone 4 have significantly shorter mean VOTs than in Tone 3. As for Hakka stops, the existence of non-words does not have a significant impact. Post hoc tests show that aspirated stops in Tones 4 and 8 have significantly shorter VOT values than stops in other tones. Hakka words in Tones 4 and 8 have similar phonetic characteristics, which are short, rapid and ended by a stop. This may explain why Hakka stops in Tones 4 and 8 are shorter than stops in other tones. The results in this study indicate that lexical tone has significant influence. Therefore, it is suggested that future studies should take the effects of lexical tone into consideration in creating the stimulus words of tonal languages when analyzing the VOT values for stops, in order to reduce the risk of introducing experimental errors. However, in what way tone will affect the VOT values for stops, further studies are needed.

References

[1] H. Liu, M. L. Ng, M. Wan, S. Wang, and Y. Zhang, "The effect of tonal changes on voice onset time in Mandarin esophageal speech," *Journal of Voice*, 2008, vol. 22, no. 2, pp. 210-218.

[2] L. Lisker and A. S. Abramson, "A cross-language study of voicing in initial stops: acoustical measurements," *Word*, 1964, vol. 20, pp. 384-422.

[3] B. L. Rochet and Y. Fei, "Effect of consonant and vowel context on Mandarin Chinese VOT: production and perception," *Canadian Acoustics*, 1991, vol. 19, no. 4, pp. 105-106.

[4] T. Cho and P. Ladefoged, "Variation and universals in VOT: evidence from 18 languages," *Journal of Phonetics*, 1999, vol. 27, pp. 207–229.

[5] M. Gósy, "The VOT of the Hungarian voiceless plosives in words and in spontaneous Speech," *International Journal of Speech Technology*, 2001, vol. 4, pp. 75–85.

[6] X.-R. Zheng and Y.-H. Li, "A contrastive study of VOT of English and Korean stops," *Journal of Yanbian University*, 2005, vol. 38, no.4, pp. 99-102.

[7] T. J. Riney, N. Takagi, K. Ota, and Y. Uchida, "The intermediate degree of VOT in Japanese initial voiceless stops," *Journal of Phonetics*, 2007, vol. 35, pp. 439-443.

[8] L. Jäncke, "Variability and duration of voice onset time and phonation in stuttering and nonstuttering adults," *Journal of Fluency Disorders*, 1994, vol. 19, no. 1, pp. 21-37.

[9] P. Auzou, C. Ozsancak, R. J. Morris, M. Jan, F. Eustache, and D. Hannequin, "Voice onset time in aphasia, apraxia of speech and dysarthria: a review," *Clinical Linguistics & Phonetics*, 2000, vol. 14, no. 2, pp. 131-150.
[10] P. A. Keating, W. Linker, and M. Huffman, "Patterns in allophone distribution for voiced and voiceless stops," *Journal of Phonetics*, 1983, vol. 11, pp. 277–90.
[11] T. J. Riney and N. Takagi, "Global foreign accent and voice onset time among Japanese EFL speakers," *Language Learning*, 1999, vol, 49, no. 2, pp. 275–302.
[12] S. J. Liao, "Interlanguage production of English stop consonants: A VOT analysis," M. A. thesis, National Kaohsiung Normal University, Kaohsiung, Taiwan, 2005.
[13] B. S. Rosner, L. E. López- Bascuas, J. E. García-Albea, and R. P. Fahey, "Voice-onset times for Castilian Spanish initial stops," *Journal of Phonetics*, 2000, vol. 28, pp. 217-224.
[14] L.-M. Chen, K.-Y. Chao, and J.-F., Peng, "VOT productions of word-initial stops in Mandarin and English: a cross-language study," *Proceedings of the 19th Conference on Computational Linguistics and Speech Processing*, 2007, pp. 303-317.
[15] Y. R. Chao, Mandarin Primer, Cambridge: Harvard University Press. 1967.
[16] S.-S. He, "A contrastive study of Taiwan Hakka and Mandarin phoneme," In G.-S. Gu,（Ed.）. Introduction to Taiwan Hakka, pp. 163-192, Taipei: Wu-Nan Book Inc. 2005.
[17] C.-C. Cheng, A Synchronic Phonology of Mandarin Chinese, The Hague: Mouton. 1973.
[18] R.-F. Chung, An introduction to Taiwan Hakka phonology, Taipei: Wu-Nan Book Inc. 2004.
[19] P. Boersma and D. Weenink, Praat: doing phonetics by computer (Version 5.1.12), 2009. [Computer program]. Available: http://www.praat.org/.