

Tonal Effects on Voice Onset Time

Jui-Feng Peng*, Li-mei Chen*, and Chia-Cheng Lee*

Abstract

This study examines the influence of lexical tone on voice onset time (VOT) in Mandarin and Hakka spoken in Taiwan. The 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 significant influence on the VOT values for stops. The results are important as they suggest that future studies should take the influence of lexical tone into account when studying VOT values and when designing wordlists for stops in tonal languages. In Mandarin, stops' VOT values, from the longest to the shortest, are in MR, FR, HL, and HF tones. This sequence is the same as in Liu, Ng, Wan, Wang, and Zhang (2008). Later, however, it was found that it is very likely that the sequence results from the existence of non-words. In order to produce non-words correctly, participants tended to pronounce them at a slower speed, especially those in MR tone. Therefore, we further examined the data without non-words, in which no clear sequence was found. For Hakka, post-hoc tests (Scheffe) show that aspirated stops in entering tones, which are syllables ending with a stop, have significantly shorter VOT values than they have in other tones. Although the tonal effects on VOT values are not consistently found in different sets of data, probably due to a methodology problem, the possibility of tonal effect on VOT values could not be excluded. Tonal effect, thus, should be taken into consideration in designing word lists for VOT studies. Moreover, further studies should include both real words and non-words in separate sets of word lists to verify the current study results.

Keywords: Voice Onset Time, Hakka Stops, Mandarin Stops, Tonal Effect

* Department of Foreign Languages and Literature, National Cheng Kung University, 1 University Rd., Tainan, Taiwan. Telephone: (06)2757575 ext. 52231

E-mail: leemay@mail.ncku.edu.tw

The author for correspondence is Li-mei Chen.

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 one of the reliable acoustic features for differentiating consonant stops (Cho & Ladefoged, 1999; Gósy, 2001; Lisker & Abramson, 1964; Riney, Takagi, Ota, & Uchida, 2007; Rochet & Fei, 1991; Zheng & Li, 2005) and it has been applied recently to the study of the language production of patients with language deficits or disorders (Auzou, Ozsancak, Morris, Jan, Eustache, & Hannequin, 2000; Jäncke, 1994). Among the languages being investigated, some are tonal languages, *i.e.* Mandarin, Cantonese, and Taiwanese. In a tonal language, the duration of each lexical tone (which can change the meaning of a word) differs slightly. Consequently, it is possible that lexical tone will affect stops' VOT to some extent; nevertheless, few studies have taken this factor into consideration when studying tonal languages. Therefore, the current study examines two tonal languages, Mandarin and Hakka spoken in Taiwan, to verify the effects of lexical tone. It is hoped that the results of the current study can establish the groundwork for future studies related to VOTs in tonal languages. If lexical tone does influence VOT, it should be considered when creating speech materials in future studies for tonal languages.

2. Literature Review

2.1 Voice Onset Time (VOT)

Lisker and Abramson (1964) defined VOT as the temporal interval from the release of an initial stop to the onset of glottal pulsing for a following vowel. VOT has been considered a reliable phonetic cue for categorizing stop consonants (*i.e.*, voiced versus voiceless or unaspirated versus aspirated) in various languages (Cho & Ladefoged, 1999; Gósy, 2001; Keating, Linker, & Huffman, 1983; Lisker & Abramson, 1964; Riney *et al.*, 2007; Rochet & Fei, 1991; Zheng & Li, 2005). In addition, by comparing VOT values for stops produced by native and non-native speakers for specific languages, researchers have put forth specific suggestions for language learning and teaching (Liao, 2005; Riney & Takagi, 1999; Zheng & Li, 2005). Moreover, recently, researchers have studied production deficits of aphasia, apraxia, and stuttering patients by observing their VOT values for stops (Jäncke, 1994; Auzou *et al.*, 2000; Tsen, 1994).

2.2 Factors Affecting Voice Onset Time

When investigating stops, researchers found that the VOT values for stops varied in relation to the place of articulation. Lisker and Abramson (1964) demonstrated that, for both unaspirated and aspirated stops, velar stops have longer mean VOT values than alveolar and bilabial stops.

In the languages examined, except for Tamil, Cantonese, and Eastern Armenian, alveolar stops tend to have longer mean VOT values than bilabial stops. Cho and Ladefoged's (1999) study further revealed that velar stops have the longest mean VOT values, alveolar stops have intermediate mean VOTs, and bilabial stops have the shortest mean VOTs, with the exception of Navajo and Dahalo. The fact that VOT values get longer when the place of articulation moves from an anterior to a posterior position is confirmed in most languages (Cho & Ladefoged, 1999; Lisker & Abramson, 1964; Rosner, López-Bascuas, García-Albea, & Fahey, 2000; Zheng & Li, 2005); nevertheless, some exceptions exist, including Hungarian, Japanese, and Mandarin.

As for the influence of vowel context, Lisker and Abramson (1967) reported that the vowels following the consonants do not have a significant effect on stops' VOTs. Recently, however, other researchers have made opposing claims. Morris, McCrea, & Herring (2008), who studied English word-initial stops, claimed that stops preceding the high vowels /i/ and /u/ had longer VOTs than stops preceding the low vowel /a/. Similar results were revealed in Rochet and Fei (1991), Chao, Khattab, and Chen (2006), and Chen, Chao, and Peng (2007) studies of Mandarin and Gósy's (2001) study of Hungarian. Furthermore, Gósy (2001) indicated that the higher the tongue position, the longer the VOTs for the preceding voiceless stops. Fant (1973), however, found the opposite to be true in a study of Swedish: the VOTs for aspirated stops preceding /a/ were longer than the VOTs for stops preceding /i/ and /u/. Fant's results are extraordinary, as most studies report that stops preceding high vowels tend to have longer VOTs than stops preceding low vowels.

Moreover, speaking rate might have influences on stops' VOTs. Kessinger and Blumstein (1997), who investigated English, French, and Thai, claimed that the speaking rate affected VOT values for long lag stops in Thai and English and for pre-voiced stops in Thai and French, but did not influence VOTs in the short lag category. Magloire and Green (1999) suggested that the speaking rate affected English monolinguals' VOT production and Spanish monolinguals' production of pre-voicing of the voiced stops. By examining English, Kessinger and Blumstein (1998) also reported that both VOT and vowel duration increased as the speaking rate slowed down. Gósy's (2001) study results further proved this. Gósy found that Hungarian bilabial and velar stops had significantly shorter mean VOTs in natural fluent speech than in carefully produced speech. Therefore, it is reasonable to expect that, in careful speech, the speaking rate will decrease and the accompanying VOT will get longer.

The VOT values for word-initial stops in various languages have been extensively investigated. Although some of the languages studied are tonal languages (*e.g.*, Mandarin, Taiwanese, and Cantonese), few studies have considered the effects of lexical tone when designing speech materials (Chao *et al.*, 2006; Chen *et al.*, 2007; Liao, 2005; Lisker & Abramson, 1964; Rochet & Fei, 1991). Gu (2005) claimed that tone is affected primarily by

pitch. Different tones have different pitch levels, which are determined by the vibrating frequency of the vocal cord. When the vocal cord tenses, the frequency of vibration increases, resulting in a higher pitch level. Conversely, the pitch level is low when the vocal cord is loose. Liu, Ng, Wan, Wang, and Zhang (2008) speculated that VOT durations may be affected by tone, as different tones have different fundamental frequencies and pitch levels, which are determined primarily by the tension of the vibrating structure. In order to achieve different levels of tension, different amounts of time might be needed. Consequently, VOT values may vary when they occur in different lexical tones. Gu (2005) further indicated that duration affects lexical tone to some extent; for example, in Hakka, the entering tone is short and rapid, meaning less time is needed to produce it. In a tonal language, the durations for each lexical tone are slightly different; therefore, it is reasonable that lexical tone might have some effects on stops' VOTs. Liu *et al.* (2008), who studied the effect of tonal changes on VOTs between normal laryngeal and superior esophageal speakers of Mandarin Chinese, reported an important finding. Normal laryngeal speakers produce significant differences in VOT values as a result of lexical tones. According to their results (Figure 1), stops in the High-falling tone have significantly shorter mean VOT values than stops in the Mid-rising tone and Falling-rising tone. Nevertheless, it should be noted that, in Liu *et al.*'s study, some of the speech materials were non-words. The researchers did not determine whether participants produced real words and non-words differently; therefore, more studies examining the influences of tone are needed. By carrying out a systematic study with respect to the influence of lexical tone on a stop's VOT using two tonal languages (*i.e.*, Mandarin and Hakka spoken in Taiwan), the current study aims to create a foundation for future linguistic studies focused on tonal languages.

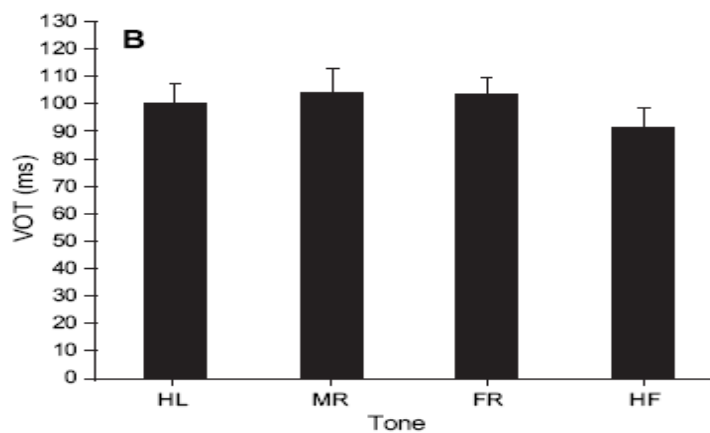


Figure 1. VOTs for Mandarin stops in individual tones produced by normal laryngeal speakers (Taken from Liu *et al.*, 2008).

2.3 Tones in Mandarin and Hakka Spoken in Taiwan

Mandarin and Hakka only have voiceless stops; therefore, the current study investigates the unaspirated stops /p, t, k/ and aspirated stops /p^h, t^h, k^h/. In addition, Mandarin and Hakka are tonal languages, in which a word's meaning can be changed by the tone in which it is pronounced. Chao (1967) 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 a Mid-rising tone in Mandarin is 35, which indicates that the pitch will go from middle to high. Mandarin has four contrasting lexical tones: High-level (HL) (55), Mid-rising (MR) (35), Falling-rising (FR) (214), and High-falling (HF) (51). Sixian Hakka has six contrasting lexical tones: low-rising (LR) (24), mid-falling (MF) (31), high-level (HL) (55), low-entering (LE) (32), low-level (LL) (11), and high-entering (HE) (55). Among them, LE and HE tones are short and rapid, and the words in these two tones end in a stop, like /p/, /t/, /k/.

Mandarin Chinese and Hakka have specific tone sandhi rules. In Mandarin, FR tone, which has the longest duration among the four lexical tones, becomes MR tone when followed by another FR tone (Cheng, 1973). In Sixian Hakka, LR tone becomes LL tone when preceding a LR tone, HL tone, or HE tone. Therefore, tone sandhi rules are taken into consideration when developing speech materials in order to avoid the combinations that might cause tonal change.

Mandarin (Chao, 1967)

FR Tone → MR Tone / _____ {FR Tone}

Sixian Hakka (Chung, 2004)

LR Tone → LL Tone / _____ {LR Tone, HL Tone, HE Tone}

3. Methodology

This study examined word-initial unaspirated stops /p, t, k/, and aspirated stops /p^h, t^h, k^h/, in combination with three corner vowels /i, u, a/ in Mandarin and Hakka spoken in Taiwan. Except for participants and speech materials, the methodology employed for both languages was the same.

3.1 Participants

In this study, the Mandarin and Hakka participants were different. The Mandarin participants included 15 male and 15 female Mandarin speakers from Tainan City with an age range from 23 to 33 years (mean = 27.2 years). All participants had grown up in Taiwan and had no hearing or speech defects. For Hakka, Sixian Hakka was chosen because it is the most extensively used Hakka dialect in Taiwan. The average age of the 21 participants - 11 men and 10 women - was 51 years, with the oldest being 80 and the youngest being 36. All of the participants for Hakka were also fluent Mandarin speakers as Mandarin is the official language in Taiwan. In the current study, the age range of Mandarin participants is controlled to within 10 years to avoid the effect of age difference. As for Hakka participants, the age-range was quite wide because it is not easy to find fluent Hakka speakers.

3.2 Data Collection

The speech materials in both languages were combinations of six stops /p, t, k, p^h, t^h, k^h/ and three vowels /i, u, a/, resulting in 18 combinations. Mandarin's 4 contrasting lexical tones meant that a total of 72 monosyllabic words were created; among them, 18 combinations do not have corresponding Chinese characters in Mandarin. The 6 contrasting lexical tones in Sixian Hakka resulted in 108 monosyllabic words, 12 of which do not have corresponding Chinese characters. Chen *et al.* (2007) 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 of the words were followed by another word in order to create meaningful disyllables, including non-words. For example, the Mandarin word /pi/ was followed by another word /p^huo/ to become the existing disyllable /pi p^huo/ "force." Even non-words were arranged in disyllabic forms to give them a more natural-like quality. Since the neutral tone in Mandarin never occurs in phrase-initial position, it was not evaluated in this study. The structure of non-words was the same as real words, which is a CV syllable with one consonant (stop) and one vowel (corner vowel). For example, there is no /k^ha/ in MR tone in Mandarin or /pu/ in MF tone in Hakka, so non-words were created for these combinations. The way we measure VOTs of non-words is the same as with the real words. For example, /k^ha/ is measured from target consonant /k^h/ to /a/.

The corpus was arranged randomly. Participants were asked to read the words out loud in a normal voice and at a comfortable rate. After finishing, the participants were asked to read the words a second time. Therefore, two groups of data were gathered for each participant. All speech was recorded using a 24 bit WAVE/MPS recorder, connected to AKG C520 Head-Worn Condenser Microphone positioned approximately 10 to 15 centimeters from the participant's mouth in a quiet room.

3.3 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 2. The values of both the waveform and spectrogram were recorded, but the VOTs were determined primarily through waveform analysis, with the values in the spectrogram being 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.

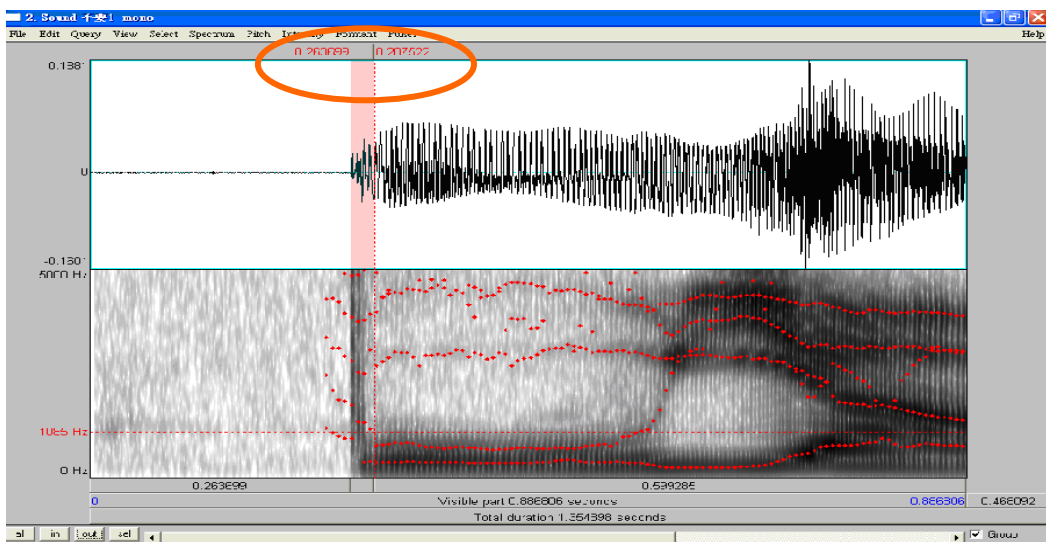


Figure 2. Spectrogram and waveform for Mandarin word /pu iau/ ‘don’t want’.
The values in the circle are the starting and endpoints of VOT.

The VOT values were measured by one investigator. Furthermore, 10% of each recording (selected randomly) was re-measured by another investigator to verify the reliability of the results. Ultimately, 7 Mandarin words and 11 Hakka words for each recording were re-measured. Pearson’s product-moment correlations (Gravetter & Wallnau, 2008) indicated high inter-rater agreement for both Mandarin and Hakka data (Mandarin: $r = .995$, $p < .001$; Hakka: $r = .978$, $p < .001$).

When analyzing the data, VOT values for mispronounced words were omitted. Moreover, data for Hakka /pi/ in HE Tone were not analyzed due to incorrect word choices. A four-way mixed factorial ANOVA (Montgomery, 2009; the same test was used in Francis, Ciocca, & Yu, 2003) (place of articulation by vowel context by lexical tone by gender) was used to examine whether the variables significantly influenced each stop’s VOT. In addition, differences between the examined targets were analyzed using T-test or post-hoc tests

(Scheffe) (Gravetter & Wallnau, 2008); results were considered significant when the p value was less than 0.05.

Four-way mixed factorial ANOVA can be illustrated in the following formula (Montgomery, 2009).

$$y_{ijklm} = \mu + \tau_i + \beta_j + \gamma_k + \sigma_l + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\tau\sigma)_{il} + (\beta\gamma)_{jk} + (\beta\sigma)_{jl} + (\gamma\sigma)_{kl} + (\tau\beta\gamma)_{ijk} + (\tau\beta\sigma)_{ijl} + (\beta\gamma\sigma)_{jkl} + (\tau\gamma\sigma)_{ikl} + (\tau\beta\gamma\sigma)_{ijkl} + \varepsilon_{ijklm}$$

$i = 1, 2, 3$ place

$j = 1, 2, 3$ vowel

$k = 1, 2, 3, 4$ tone

$l = 1, 2$ gender

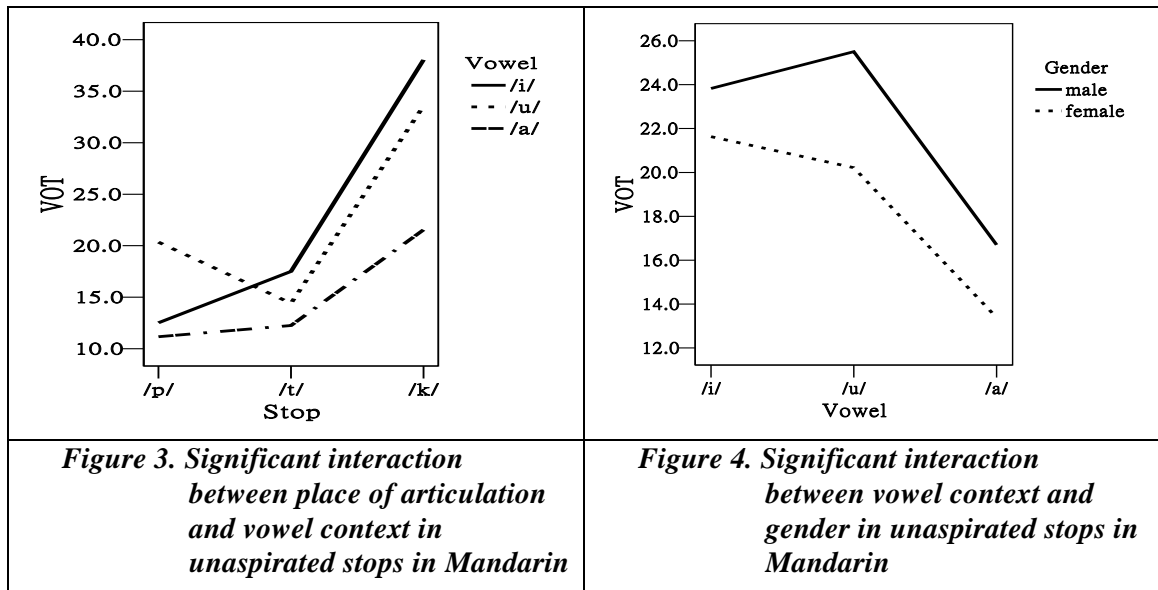
$m = 1, 2, \dots, 30$ subjects

4. Results and Discussion

When examining the VOT values for Mandarin stops, it became apparent that they tend to be longer than the mean VOT values reported in the studies by Liao (2005), Chao *et al.* (2006), and Chen *et al.* (2007) and shorter than those reported by Rochet and Fei (1991). Examining the methodologies in these previous studies indicated that the speech materials in Rochet and Fei's study were monosyllabic, but disyllabic in the remaining four studies. Gósy (2001) claimed that speakers speak in a careful and disciplined way while uttering syllables and words in isolation. Therefore, participants are expected to produce monosyllables in a more careful manner as their speech tempo decreases. According to Kessinger and Blumstein (1998), VOTs get longer while the speaking rate slows down. This may explain why the mean VOTs for Mandarin stops in Rochet and Fei's study were longer than in other reports. Another possible explanation might be regional differences in the target language. Although the target language is the same, participants in Rochet and Fei's study grew up in Mainland China, while participants in the other studies were raised in Taiwan. Consequently, regional differences might be another origin of the variations. As a result, comparing the results of Liao (2005), Chao *et al.* (2006), and Chen *et al.* (2007) to those of the present study indicates that the mean VOT values for the stops in the present study tend to be longer than their counterparts in the other studies. Such differences stem primarily from the existence of non-words in the current study. During the recording process, speakers tended to produce non-words more carefully and at a lower speed because they were not familiar with the non-words. An examination of the data both with and without non-words separately demonstrated that the existence of non-words resulted in stops having longer mean VOTs.

4.1 Mandarin Chinese

The statistical analyses were conducted using a four-way mixed factorial ANOVA. For Mandarin unaspirated stops, the results showed a primary effect of place of articulation, $F(2, 972) = 522.9680$; vowel context, $F(2, 972) = 117.3569$; lexical tone, $F(3, 972) = 6.5506$; and gender $F(1, 972) = 56.9180$ (all $p < .001$). These results indicate that the stop place of articulation (bilabial, alveolar, and velar), vowel context (/i/, /u/, /a/), lexical tone (HL, MR, FR, HF), and gender do create significant differences in VOT values of word-initial unaspirated stops. Furthermore, significant two-way interactions were also observed (Figures 3-4) between place of articulation and vowel context and between vowel context and gender. A complete ANOVA table showing interaction between variables is listed in Appendix 1.



As for aspirated stops, the results demonstrated a main effect of place of articulation, $F(2, 972) = 95.2742$; vowel context, $F(2, 972) = 43.5079$; lexical tone, $F(3, 972) = 12.0121$; and gender $F(1, 972) = 20.3186$, thereby indicating that VOT values of word-initial aspirated stops would vary in accordance with place of articulation, vowel context, lexical tone and gender (all $p < .001$). Significant two-way interactions occurred between place of articulation and vowel context, between place of articulation and lexical tone, and between vowel context and gender (Figures 5-7). For both unaspirated and aspirated stops in Mandarin, no significant three-way and four-way interactions were evident between variables. A complete ANOVA table showing interaction between variables is listed in Appendix 2.

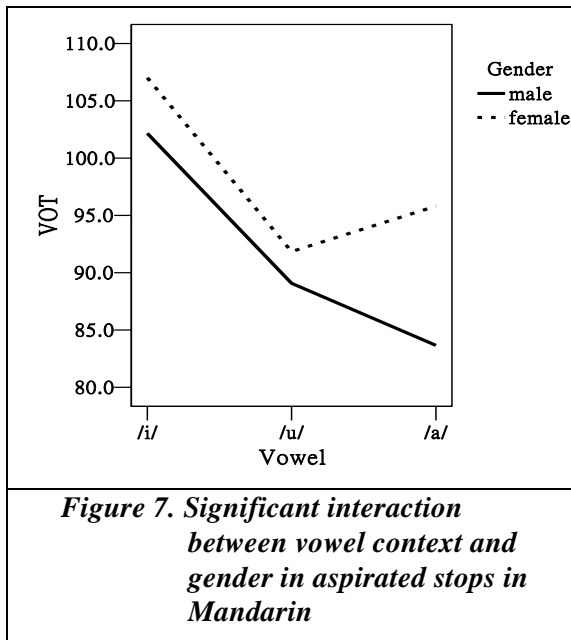
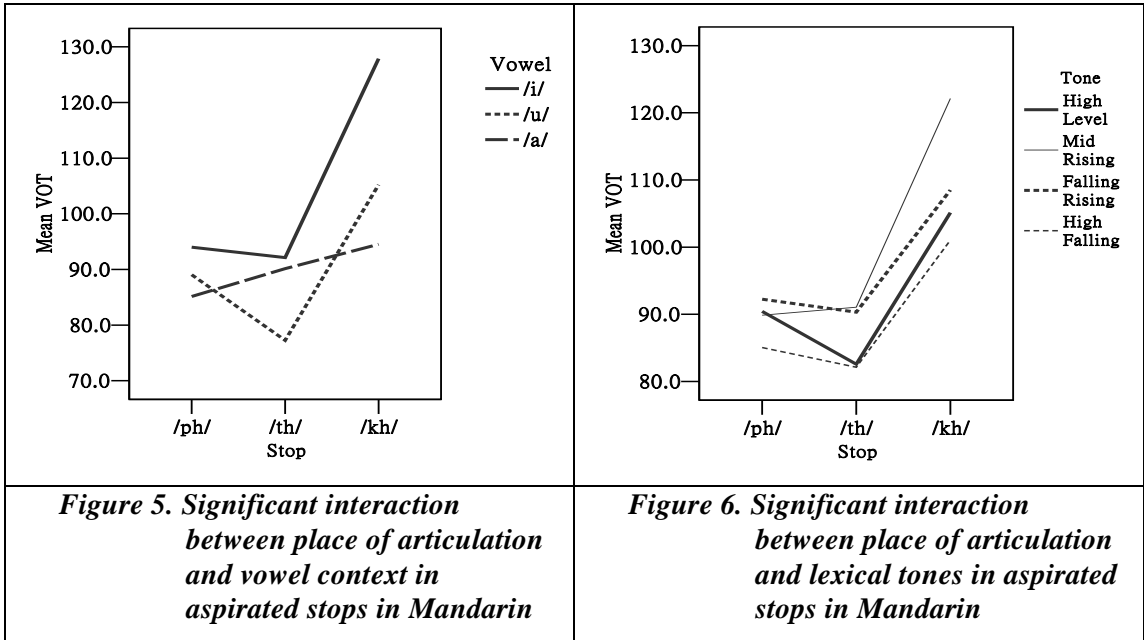


Table 1 lists mean VOT values and standard deviation of Mandarin stops in each lexical tone. ANOVA tests revealed that the lexical tone significantly influences VOTs of stops [unaspirated stops, $F(3, 972) = 6.5506, p < .001$; aspirated stops, $F(3, 972) = 12.0121, p < .001$]. A post-hoc test reveals that aspirated stops in HF have significantly shorter mean

VOTs than stops in MR and FR (all $p < .05$). In addition, for both unaspirated and aspirated stops, stops in MR have the longest mean VOTs while stops in HF have the shortest mean VOTs. VOTs, from longest to shortest, occurred in MR, FR, HL, and HF - the same sequence as in Pearce (2009) and Liu *et al.* (2008). Yet, it is worth noting that, in both studies, some of the speech materials were non-words. It was subsequently determined 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 slower speed, making the VOTs longer.

Therefore, the current study further examined the data without non-words, in which the main variation occurred in participants' productions of stops in MR tone. When analyzing the data with non-words, Mandarin unaspirated and aspirated stops in MR tone had the longest mean VOTs. Nevertheless, when excluding non-words, the results revealed that stops in MR tone did not have the longest mean VOTs, and unaspirated stops in MR tone even had significantly shorter mean VOTs than in HL and FR tones. The divergence revealed that participants' productions of real words or non-words in MR tone were quite different. In addition, ANOVA tests revealed that lexical tone does not significantly influence stops' VOTs by analyzing the data without non-words. Further studies are needed to have separate sets of wordlists for real words and non-words to verify the current findings.

Table 1. Mean VOT values of Mandarin stops with different 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
HL	20.20	(11.90)	92.72	(25.53)	17.71	(9.95)	88.69	(20.4)
MR	21.10	(12.68)	101.02	(30.21)	13.99	(6.03)	89.47	(23.31)
FR	20.89	(13.35)	97.03	(27.75)	17.00	(10.98)	92.30	(23.49)
HF	18.42	(9.94)	89.4	(25.72)	16.32	(9.07)	85.62	(24.18)

Table 2. Mean VOT values of six Mandarin stops with different lexical tones. All measurements are in milliseconds (ms).

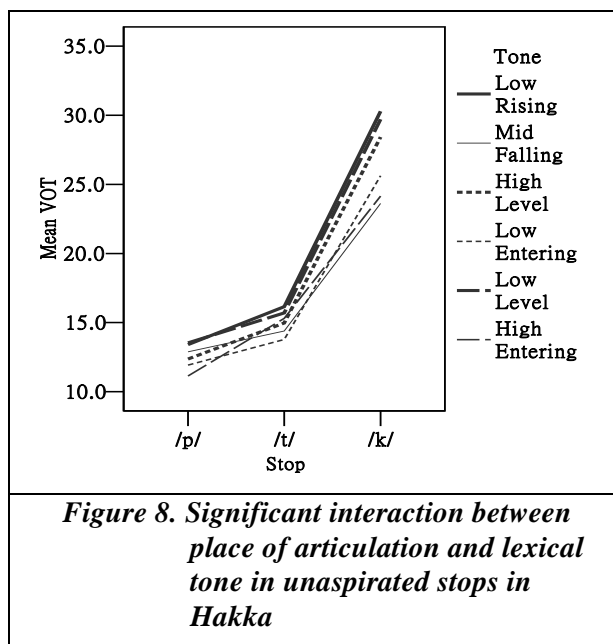
	With non-words (mean)						Without non-words (mean)					
	p	t	k	p ^h	t ^h	k ^h	p	t	k	p ^h	t ^h	k ^h
HL	14.4	14.7	31.5	90.4	82.6	105.1	12.0	14.7	27.9	90.4	82.6	95.2
MR	16.0	15.1	32.2	89.9	91.1	122.1	12.3	15.1	*	89.9	88.9	97.4
FR	15.0	14.9	32.7	92.3	90.3	108.5	11.3	14.9	34.7	90.2	90.3	97.8
HF	13.3	14.2	27.7	85.0	82.1	101.0	11.8	14.2	31.8	85.0	82.1	*

*No real word data

In the data with non-words in Table 2, individual stops mostly follow the pattern of MR with the longest VOT and HF with the shortest VOT, especially for /p, t, t^h, k^h/. As for the data without non-words, the general pattern is also found in /p, t/.

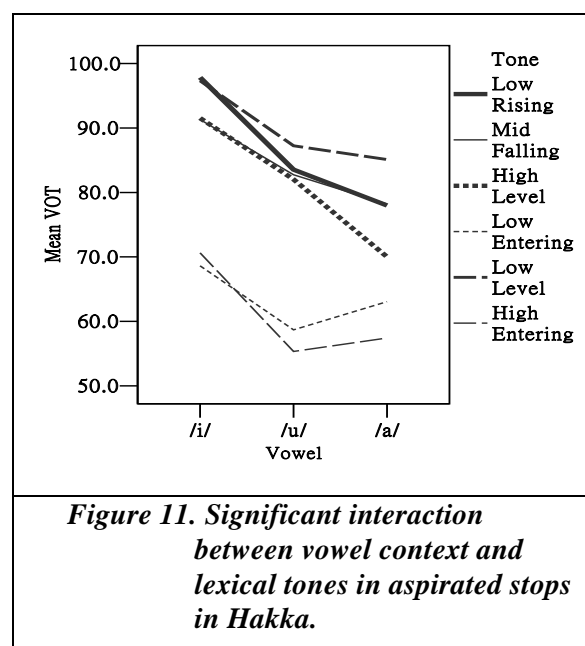
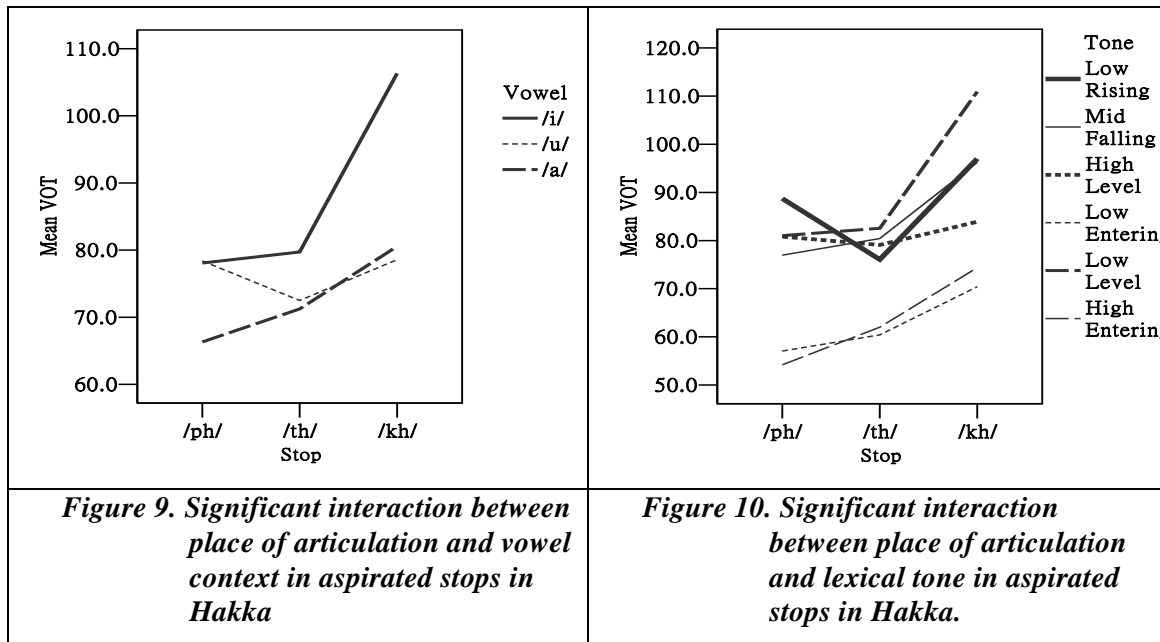
4.2 Lexical Tone and VOT in Hakka

For Hakka unaspirated stops, the results showed a main effect of place of articulation, $F(2, 843) = 404.3395$; vowel context, $F(2, 843) = 69.8958$; lexical tone, $F(5, 843) = 6.3054$; and gender $F(1, 843) = 34.2724$ (all $p < .001$). Similar to the findings in Mandarin stops, these results indicate that stop place of articulation (bilabial, alveolar, and velar), vowel context (/i/, /u/, /a/), lexical tone (HL, MR, FR, HF), and gender do make significant differences in VOT values of word-initial unaspirated stops. Significant two-way interactions occurred between the place of articulation and lexical tone (Figure 8), and significant three-way interactions occurred among place of articulation, vowel context, and lexical tone. A complete ANOVA table showing interaction between variables is listed in Appendix 3.



As for aspirated stops, the results also demonstrated a main effect of place of articulation, $F(2, 798) = 55.6543$; vowel context, $F(2, 798) = 44.7708$; lexical tone, $F(5, 798) = 46.4587$; and gender $F(1, 798) = 42.0266$, thereby indicating that VOT values of word-initial aspirated stops would vary in accordance with place of articulation, vowel context, lexical tone, and gender (all $p < .001$). Significant two-way interactions existed between place of articulation and vowel context, between place of articulation and lexical tone, and between vowel context and lexical tone (Figures 9-11). In addition, significant three-way interactions occurred among

place of articulation, vowel context, and lexical tone. A complete ANOVA table showing interaction between variables is listed in Appendix 4.



The mean VOT values and standard deviations for Hakka stops in each lexical tone are shown in Table 3. ANOVA tests indicated that the lexical tone significantly influences stops'

VOTs [unaspirated stops, $F(5, 843) = 6.3054, p < .001$; aspirated stops, $F(5, 798) = 46.4587, p < .001$]. Unaspirated and aspirated stops in LR and LL have longer mean VOTs than stops in other tones, whereas the shortest mean VOTs for both unaspirated and aspirated stops are in HE. The post-hoc test revealed that aspirated stops in HE and LE have significantly shorter mean VOTs than those in other tones (all $p < .001$). In Hakka, HE and LE tones are entering tones, which are short, rapid, and end in a stop like /p, t, k/. Gu (2005) further claimed that the durations for entering tones are shorter than the durations for other tones. Therefore, the VOTs for stops in the entering tones are shorter than stops in other tones.

Table 3. Mean VOT values of Hakka stops with different lexical tones. All measurements are in milliseconds (ms)

	Unaspirated stops		Aspirated stops	
	mean	(SD)	mean	(SD)
LR	20	(11.56)	86.83	(25.8)
MF	16.94	(8)	84.67	(26.56)
HL	18.88	(11.02)	81.32	(23.73)
LE	17.19	(9.44)	62.93	(18.36)
LL	19.4	(11.43)	90.08	(27.08)
HE	16.11	(7.98)	61.53	(20.36)

Table 4. Mean VOT values of six Hakka stops with different lexical tones. All measurements are in milliseconds (ms)

	Without non-words (mean)					
	p	t	k	p ^h	t ^h	k ^h
LR	13.4	16.2	30.3	88.7	76.1	97.0
MF	12.9	14.4	23.6	77.0	80.4	96.2
HL	12.4	15.0	28.4	80.8	79.1	83.9
LE	11.9	13.8	25.6	57.0	60.4	70.4
LL	13.5	15.7	29.7	81.0	82.5	110.9
HE	11.1	15.3	24.2	54.2	62.0	74.4

Table 4 shows that almost all of the stops follow the general pattern where LE and HE are among the shortest.

5. Conclusion

The study results revealed that lexical tones significantly influence VOT values for stops in Mandarin (with both real words and non-words), and significant tonal effect was also found in

Hakka data of real words. Nevertheless, there is no significant tonal effect on VOT in Mandarin data with only real words.

The study results are important as they suggest that future studies should take the influence of lexical tones into account when studying VOT values and when designing wordlists for stops in tonal languages. Although the tonal effects on VOT values are not consistently found in different sets of data, probably due to a methodology problem, the possibility of tonal effect on VOT values could not be excluded. Several factors might contribute to this inconsistency. First, we used different methods in eliciting non-words production in these two languages. In Mandarin, we used Zhuyin Fuhao to guide non-words productions, which might force participants to take a few seconds to figure out the new combinations of Chinese phonetic symbols. In contrast, we asked participants to read a real word first as a clue in producing a target non-word in Hakka. Second, many of the participants in Hakka were at the age range of 50-80, and we found the participants were not flexible enough to comprehend our instructions in producing non-words. Therefore, we decided not to take the data of non-words (mostly by guessing with uncertainly) into analysis in order to have a reliable result. (The discarded cells (non-words) are less than 25% of the total, which fulfills the requirements of ANOVA test for reliable test results (Gravetter & Wallnau, 2008). Future studies should keep the method of elicitation consistent across languages, recruit participants of similar age range, and include both real words and non-words in separate sets of word lists in order to verify the current study findings. Moreover, although the results of this study indicate that lexical tones influence VOT of stops to some extent, we cannot exclude the possibility of correlation between VOT and the duration of lexical tones. Further study can explore this possibility by adopting mean durations of each tone in Mandarin as the normalized parameters upon calculating mean VOTs of stops.

References

- Auzou, P., Ozsancak, C., Morris, R. J., Jan, M., Eustache, F., & Hannequin, D. (2000). Voice onset time in aphasia, apraxia of speech and dysarthria: a review. *Clinical Linguistics & Phonetics*, 14(2), 131-150.
- Chao, K.-Y., Khattab, G., & Chen, L.- M. (2006). Comparison of VOT patterns in Mandarin Chinese and in English. *Proceedings of the 4th Annual Hawaii International Conference on Arts and Humanities*, 840-859.
- Chao, Y. R. (1967). *Mandarin Primer*. Cambridge: Harvard University Press.
- Chen, L.- M., Chao, K.- Y., & Peng, J.- F. (2007). 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*, 303-317.
- Cheng, C.- C. (1973). *A Synchronic Phonology of Mandarin Chinese*. The Hague: Mouton.

- Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, 27, 207-229.
- Chung, R.- F. (2004). *An introduction to Taiwan Hakka phonology*. Taipei: Wu-Nan Book.
- Fant, G. (1973). *Speech sounds and features*. Cambridge: MIT Press.
- Francis, A. L., Ciocca, V., & Yu, J. M. C. (2003). Accuracy and variability of acoustic measures of voicing onset. *Journal of the Acoustical Society of America*, 113(2), 1025-1032.
- Gósy, M. (2001). The VOT of the Hungarian voiceless plosives in words and in spontaneous speech. *International Journal of Speech Technology*, 4, 75-85.
- Gravetter, F. J., & Wallnau, L. B. (2008). *Statistics for the behavioral sciences* (8th ed.). Belmont, California: Wadsworth.
- Gu, G.- S. (2005). Phonology of Taiwan Hakka. In Gu, G.-S. (Ed.), *Introduction to Taiwan Hakka*, 117-161. Taipei: Wu-Nan.
- Jäncke, L. (1994). Variability and duration of voice onset time and phonation in stuttering and nonstuttering adults. *Journal of Fluency Disorders*, 19(1), 21-37.
- Keating, P. A., Linker, W., & Huffman, M. (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics*, 11, 277-290.
- Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, 25, 143-168.
- Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: some implications for perception studies. *Journal of Phonetics*, 26, 117-128.
- Liao, S. J. (2005). *Interlanguage production of English stop consonants: a VOT analysis*. Unpublished master's thesis, National Kaohsiung Normal University, Kaohsiung, Taiwan.
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384-422.
- Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language Speech*, 10, 1-28.
- Liu, H., Ng, M. L., Wan, M., Wang, S., & Zhang, Y. (2008). The effect of tonal changes on voice onset time in Mandarin esophageal speech. *Journal of Voice*, 22(2), 210-218.
- Magloire, J., & Green, K. P. (1999). A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica*, 56, 158-185.
- Montgomery, D. C. (2009). *Design and analysis of experiments, student solutions manual* (7th ed.). Hoboken, N.J.: Wiley.
- Morris, R. J., McCrea, C. R., & Herring, K. D. (2008). Voice onset time differences between adult males and females: isolated syllables. *Journal of Phonetics*, 36, 308-317.
- Pearce, M. (2009). Kera tone and voicing interaction. *Lingua*, 119(6), 846-864.

- Riney, T. J., & Takagi, N. (1999). Global foreign accent and voice onset time among Japanese EFL speakers. *Language Learning*, 49(2), 275-302.
- Riney, T. J., Takagi, N., Ota, K., & Uchida, Y. (2007). The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics*, 35, 439-443.
- Rochet, B. L., & Fei, Y. (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: production and perception. *Canadian Acoustics*, 19(4), 105-106.
- Rosner, B. S., López- Bascuas, L. E., García-Albea, J. E., & Fahey, R. P. (2000). Voice-onset times for Castilian Spanish initial stops. *Journal of Phonetics*, 28, 217-224.
- Tsen, C.Y. (1994). An investigation on voice onset time and tone production of Chinese aphasia. *Bulletin of the Institute of History and Philology Academia Sinica*, 65(1), 37-79.
- Zheng, X.- R., & Li, Y.- H. (2005). A contrastive study of VOT of English and Korean stops. *Journal of Yanbian University*, 38(4), 99-102.

Appendix 1.

Four-way factorial ANOVA for Mandarin unaspirated stops

Source of Variation	SS	df	MS	F
Place of articulation	62001	2	31000	522.9680
Vowel	13913	2	6957	117.3569
Gender	3374	1	3374	56.9180
Tone	1165	3	388	6.5506
Vowel×Place	10289	4	2572	43.3945
Gender×Place	242	2	121	2.0372
Gender×Vowel	427	2	214	3.6038
Tone×Place	544	6	91	1.5286
Tone×Vowel	252	6	42	0.7081
Tone×Gender	130	3	43	0.7283
Gender×Vowel×Place	349	4	87	1.4725
Vowel×Place×Tone	1016	12	85	1.4283
Place×Tone×Gender	148	6	25	0.4153
Tone×Gender×Vowel	104	6	17	0.2910
Tone×Gender×Vowel×Place	216	12	18	0.3035
Residual	57618	972	59	

Appendix 2

Four-way factorial ANOVA for Mandarin aspirated stops

Source of Variation	SS	df	MS	F
Place of articulation	106145	2	53073	95.2742
Vowel	48472	21	24236	43.5079
Gender	11318	1	11318	20.3186
Tone	20074	3	6691	12.0121
Vowel×Place	38558	4	9640	17.3046
Gender×Place	1273	2	636	1.1422
Gender×Vowel	4216	2	2108	3.7840
Tone×Place	10302	6	1717	3.0824
Tone×Vowel	2378	6	396	0.7115
Tone×Gender	419	3	140	0.2508
Gender×Vowel×Place	3253	4	813	1.4598
Vowel×Place×Tone	5303	12	442	0.7933
Place×Tone×Gender	1880	6	313	0.5626
Tone×Gender×Vowel	552	6	92	0.1650
Tone×Gender×Vowel×Place	3451	12	288	0.5162
Residual	541453	972	557	

Appendix 3.

Four-way factorial ANOVA for Hakka unaspirated stops

Source of Variation	SS	df	MS	F
Place of articulation	38120	2	19060	404.3395
Vowel	6590	2	3295	69.8958
Gender	1616	1	1616	34.2724
Tone	1486	5	297	6.3054
Vowel×Place	4613	4	1153	24.4639
Gender×Place	61	2	31	0.6502
Gender×Vowel	50	2	25	0.5267
Tone×Place	1104	10	110	2.3427
Tone×Vowel	837	10	84	1.7750
Tone×Gender	436	5	87	1.8519
Gender×Vowel×Place	293	4	73	1.5519
Vowel×Place×Tone	2132	19	112	2.3799
Place×Tone×Gender	350	10	35	0.7415
Tone×Gender×Vowel	235	10	23	0.4982
Tone×Gender×Vowel×Place	475	19	25	0.5302
Residual	39738	843	47	

Appendix 4.

Four-way factorial ANOVA for Hakka aspirated stops

Source of Variation	SS	df	MS	F
Place of articulation	46848	2	23424	55.6543
Vowel	37687	2	18843	44.7708
Gender	17688	1	17688	42.0266
Tone	97769	5	19554	46.4587
Vowel×Place	22984	4	5746	13.6520
Gender×Place	123	2	62	0.1465
Gender×Vowel	451	2	225	0.5357
Tone×Place	17302	10	1730	4.1108
Tone×Vowel	8157	10	816	1.9381
Tone×Gender	2030	5	406	0.9649
Gender×Vowel×Place	1436	4	359	0.8532
Vowel×Place×Tone	35729	20	1786	4.2445
Place×Tone×Gender	2244	10	224	0.5331
Tone×Gender×Vowel	4006	10	401	0.9517
Tone×Gender×Vowel×Place	4705	20	235	0.5590
Residual	335866	798	421	

