VOT productions of word-initial stops in Mandarin and English: A cross-language study

Li-mei Chen*
Kuan-Yi Chao**
Jui-Feng Peng***

Department of Foreign Languages and Literature
National Cheng Kung University

* Associate Professor, leemay@mail.ncku.edu.tw
** Lecturer, chaoky@mail.ncku.edu.tw
*** Postgraduate student, tn670124@mail.tn.edu.tw

Abstract

Voice Onset Time (VOT) is considered as one of the best methods for examining the timing of voicing in stop consonants and has been applied in the study of many languages. The present study is designed to examine VOT production for phonetically voiceless stops in Mandarin and English by native Chinese speakers. Thirty-six Taiwanese Chinese speakers recruited from National Cheng Kung University participated in this study. The results indicate the following. 1) Based on the three universal categories proposed by Lisker and Abramson (1964), for phonetically voiceless stops, Mandarin and English occupy the same place along the VOT continuum. 2) The mean VOT value for the apical stop /t/ is slightly lower than the mean value for the labial stop /p/. This does not conform to the general consensus, which states that the further back the place of articulation the longer the VOT. Very similar findings were also observed in previous studies. 3) The difference between the mean VOT values of the English /p/ and /t/ produced by Chinese speakers was subtle, while it reached significance for native English speakers. This suggests that a first language could be a crucial factor in L2 production. Future studies might examine variations in L2 production both for the same persons over time and for different speakers.
Keywords: voice onset time (VOT), voiceless stops, place of articulation

1. Introduction

Voicing contrast in stops has been discussed in phonetics and phonology for the past few decades. Beginning with Lisker and Abramson (1964), in their well-known cross-language study, voice onset time (VOT) has been widely used to differentiate stop categories across languages. Since then, VOT has come to be regarded as one of the best acoustic cues for discriminating three general stop categories, especially in word-initial position. In contrast with the considerable number of studies investigating stop voicing contrast in a variety of languages, only a few have examined Mandarin word-initial stops, not to mention comparing VOT patterns in Mandarin and English. Therefore, the purpose of this present study is threefold. First, it is intended to provide information for a general VOT pattern of Mandarin word-initial stops. By analyzing VOTs in stop consonants, linguists have concluded that for most languages, VOT values get longer as the place of articulation moves backward (Lisker & Abramson, 1964; Cho & Ladefoged, 1999; Gósy, 2001). However, there are some exceptions, such as Mandarin, which does not follow the general rule (Lisker & Abramson, 1964; Cho & Ladefoged, 1999; Chao, Khattab & Chen, 2006). The second purpose is to explore the possible effects of this phenomenon. Vowel context is also examined to determine whether there is a correlation between VOT and subsequent vowels. Moreover, to date no study has focused on comparing the in-depth differences between Mandarin and English, except for Chao et al. (2006) who pinpoints the existence of subtle differences between the two languages. Thus, the third aim is to compare VOT patterns of the two languages and observe L2 production (i.e. English production) by native Chinese speakers.

2. Literature review

2.1 Voice onset time

Lisker and Abramson (1964) conducted a cross-language investigation of word-initial stops in 11 languages and define voice onset time as the temporal interval from the release burst of the stop consonant to the onset of the first formant (F1) frequency that reflects glottal vibration. Following their study, VOT has been widely used to examine voicing contrast in stops in many languages (Keating, Linker, and Huffman, 1983; Rochet & Fei, 1991; Cho and Ladefoged, 1999; Gósy, 2000; Khattab, 2000; Zheng & Li, 2005; Riney, Takagi, Ota, and Uchida, 2006). In addition to investigating phonetic characteristics of voiced and voiceless stops in various languages, some researchers have studied VOT with respect to place of articulation, speaking rate, bilingual language learners, and vowel environment (Kewley-Port, Pisoni, and Studdert-Kennedy, 1983; Port and Rotunno 1979; Kessinger and Blumstein 1997; Benkí, 2001; Kehoe, Lleó, and Rakow, 2004). Thus, VOT is one of the main acoustic cues used to measure the timing of voicing in stops.

Although VOT is now used across the world as a linguistic cue, some researchers, however, challenge its role and importance as a reliable measure for separating phonemic categories. In their study examining voicing contrast among French-English bilinguals, Caramazza, Yeni-Komshian, Zurif, and Carbone (1973) argue that voice onset time is ineffective at differentiating stop categories. Bohn and Flege (1993) also question its importance to the perception of stop voicing. Docherty (1992) indicates that VOT narrowly concentrates on word-initial stops. Moreover, Klatt (1975) even suggests five other acoustic cues that are equally important to voice onset time: that is, low frequency energy in subsequent vowels, burst loudness, fundamental frequency, pre-voicing, and segmental duration. Even if VOT does have limitations, it is still regarded as one of the most important acoustic parameters for distinguishing voicing contrast, especially for word-initial stops.

2.2 VOT category

In Lisker and Abramson's 1964 study, all stops are classified into three groups depending on the number of stop categories in each language. VOT ranges for the three stop categories are -125 to -75ms, 0 to +25ms, and +60 to +100ms. Cho and Ladefoged (1999) also provide VOT ranges for occlusives, particularly in voiceless aspirated and unaspirated stops. Rather than three categories, they distinguish four: unaspirated, slightly aspirated, aspirated, and highly aspirated. The approximate mean VOT values for each category are 30 ms, 50 ms, 90 ms, and over 90 ms, respectively. In agreement with Lisker and Abramson's (1964) categorization, on the basis of Cho and Ladefoged's (1999) categorization, stops in Mandarin and English are found to occupy the same place along the VOT continuum, whereas stops in the two languages do not belong to the same range along the continuum, especially for voiceless aspirated occlusives. Chao, Khattab, and Chen's (2006) findings confirm Cho and Ladefoge's classification and reveal that for voiceless aspirated stops, Mandarin falls into the 'highly aspirated' region while English belongs to 'highly aspirated' category. A comparison of the different stop categories in Mandarin and English is given in section 2.4, below.

2.3 Effect on VOT

2.3.1 Place of articulation

Some researchers have reported a significant link between place of articulation and voice onset time. Cho & Ladefoged (1999) propose some possible relations including 1) the further back the closure, the longer the VOT; 2) the more extended the contact area, the longer the VOT; and 3) the faster the movement of the articulator, the shorter the VOT. Of these three suggested links, the present study focuses on the first in connection with Mandarin. In addition to this first principle, it may be stated that the velar stop /k/ has the longest VOT duration and bilabial stop /p/ the shortest, with the alveolar stop /t/ in between the two (Lisker & Abramson, 1964). Factors used to explain why VOT is longer when articulation takes place nearer the back of the mouth include aerodynamics, articulatory movement velocity, and differences in the mass of the articulators (Cho & Ladefoged, 1999).

The size of the supraglottal cavity behind the constricted points should be taken into consideration when considering the impact of aerodynamics. The cavity behind the velar stop has a smaller volume than that behind the alveolar and bilabial stops. In other words, the velar stop is under greater pressure when airflow is released; therefore, it might take longer to produce a velar stop, and the VOT value for the velar stop might be longer than either the alveolar or the bilabial stop. As for articulatory movement velocity, Cho and Ladefoged (1999) claim that the tip of the tongue and the lips move faster than the back of the tongue; moreover, the tongue tip moves faster than the lower lip. This may explain why in many languages velar stops have longer VOT than labial and alveolar stops. However, articulatory movement velocity does not affect alveolar and bilabial stops in this way in all languages, which implies that other factors are involved. In reference to the extent of articulatory contact area, Cho and Ladefoged (1999: 211) claim that, "In general, stops with a more extended articulatory contact have a longer VOT." In summary, it is indubitable that velar stops have longer VOT than the two other stops. However, no final conclusion may be reached in the case of labial and alveolar stops.

Although there is general agreement that the further back the place of articulation, the longer the VOT, there are still some exceptions. Lisker and Abramson's (1964) study reports that unaspirated stops in Tamil and aspirated stops in Cantonese and Eastern Armenian do not follow this rule. It is found that the VOT of alveolar /t/ is shorter than bilabial stop /p/, but the velar stop /k/ still has the longest VOT. Studies by Rochet and Fei (1991) and Chao et al. (2006) arrive at similar results. Investigating Mandarin Chinese, they conclude that the VOT duration for /t/ does not confirm the predictions; on the contrary, it is shorter than the VOT for /p/. The cause of this phenomenon is still unknown.

2.3.2 Vowel context

How vowels influence the VOT of preceding stops is still an open question. Lisker and Abramson (1967) propose that following vowels have no significant influence on VOTs, while other researchers apply similar research methods, but more systematically, and find that VOTs are longer when followed by tense high vowels (Klatt, 1975; Weismer, 1979). Similar results are obtained in Port's (1979) study, which analyzes VOT for English word-initial stops, and in Gósy's research, which examines Hungarian voiceless plosives. Rochet & Fei (1991) also reach similar findings with respect to Mandarin stops, claiming that "the nature of the vowel had a significant effect on the VOT values of the preceding consonants" (p. 105). In other words, word-initial stops have longer VOT values when followed by either of the high vowels /i/ or /u/ than when followed by the low vowel /a/. This accords with the results presented in Chao et al's (2006) study which examines the Mandarin Chinese of Taiwanese speakers. By contrast, however, Fant (1973) finds that for Swedish aspirated stops, VOTs are longer when stops are followed by /a/ than /i/ or /u/. Although the finer points of the issue are still undecided, a general conclusion that may be made is that vowel context does have some effects on voice onset time.

2.4 Mandarin and English stops and VOT patterns

In Lisker and Abramson's (1964) study, VOT measurements occurring before the release burst are said to have negative values, called 'voicing lead', whereas 'voicing lag' refers to measurements occurring after the release burst and are assigned positive values. Following these definitions, Keating (1984) subdivides the voicing lag dimension into 'short lag' (20–35ms) and 'long lag' (over 35ms). On the basis of this classification, stops are divided into three phonetic categories: voiced, voiceless unaspirated, and voiceless aspirated. Mandarin and English are said to contain two stop categories; detailed descriptions of the stops in these two languages are elaborated in the following sections.

2.4.1 English stops

Although, as Keating (1984) mentions, English has a great deal of positional variation, in the present study only syllable initial stops are discussed. English is known to contrast voiced and voiceless phonemes in word-initial position, while voiced stops are said to have two possible phonetic realizations, voiced or voiceless unaspirated (Keating, Linker, & Huffman, 1983; Keating, 1984; Docherty, 1992). Lisker and Abramson (1964) provide two sets of VOT values for English voiced stops (/b, d, g/), one with a positive short lag, and the other with a negative voicing lead. They further suggest that only a single type of phonetic representation is produced by each native speaker. Klatt (1975) measures VOT values for English stops and reports positive values for both voiced /b, d, g/ and voiceless unaspirated stops /p, t, k/. Keating (1984) also points out that English voiced stops are sometimes pronounced with some lead values but mainly with short lag and long lag. Table 1 shows mean VOTs for English stops, as reported by Lisker and Abramson (1964), Klatt (1975), and Docherty (1992).

Table 1. Mean VOTs for English stops

	Lisker & Abramson, 1964 (AE)	Klatt, 1975	Docherty, 1992 (BE)
	Mean	**Mean**	**Mean**
/p'/	58	47	42
/t'/	70	65	64
/k'/	80	70	62
/p/		12	
/t/		23	
/k/		30	
/b/	1/-101	11	15
/d/	5/-102	17	21
/g/	21/-88	27	27

(AE=American English; BE=British English. All measurements are in milliseconds (ms). Note: /p', t', k'/ represents voiceless aspirated stops, while /p, t, k/ refers to voiceless unaspirated stops.

2.4.2 Mandarin stops

It is known that all Mandarin stops are phonetically voiceless and that aspiration is the only distinctive phonetic feature, differentiating two phonemic categories: voiceless unaspirated /p, t, k/ and voiceless aspirated /p', t', k'/. Unlike in English, stops in Mandarin occur only in word-initial position. Moreover, Mandarin stops fall into short lag versus long lag patterns.

Table 2 juxtaposes mean Mandarin VOTs, as measured by different researchers. As well as Rochet and Fei's (1991) study of Mandarin Chinese, Liao (2005) and Chao et al. (2006) focus on Taiwanese Chinese accents. Two points are of note. First, as the table shows, VOT values for Mandarin /p', t', k'/ are obviously higher than their equivalents in English. This may imply that for voiceless aspirated stops, especially for the velar /k'/, Mandarin and English may occupy different areas along the VOT continuum. Secondly, all values for /t'/ production are close to, but slightly lower than, the values for /p'/. It is interesting to note the possible effect of not conforming to the general pattern with respect to place of articulation.

Table 2. Mean VOTs in Mandarin

	Rochet & Fei, 1991 (MC)	Liao, 2005 (TC)	Chao et al., 2006 (TC)
	Mean	**Mean**	**Mean**
/p'/	99.6	75.4	82
/t'/	98.7	71.4	81
/k'/	110.3	98.8	92
/p/		17.9	14
/t/		18.6	16
/k/		28	27

(MC=Mandarin Chinese; TC=Taiwanese Chinese accent. All measurements are in milliseconds (ms). Note: Rochet & Fei only provide the mean VOT for voiceless aspirated /p', t', k'/.

3. Methodology

3.1 Aims of the experiment

As mentioned above, some studies have examined VOT in Mandarin Chinese, but few have attempted to compare Mandarin and English VOT patterns, particularly with respect to voiceless aspirated stops. To the best of our knowledge, so far only Chao et al. (2006) have compared VOT patterns in these two langauges, and they found that there are indeed subtle differences in VOT production between Mandarin and English. Therefore, the aim of the present experiment is to compare Mandarin and English VOT patterns.

3.2 Stimuli

It is known that, in Mandarin, stops occur only in the word-initial position; moreover, all stops are phonetically voiceless and they are only distinguished by aspiration. The present experiment examines only voiceless stops in the initial position. Klatt (1975) finds that the differences in VOT values relate to the environment of the following vowel. Therefore, in this experiment each of the stops is augmented by three peripheral vowels; that is, two high vowels, /i/ and /u/, and one low vowel, /a/. The Mandarin word list consists of 16 words (excluding /k'i/ and /ki/, as no meaningful lexical items for /k'i/ and /ki/ exist in Chinese). Note that compound words (two or three characters side by side forming a 'word') are used rather than single characters because they are more complete and more sense to the subjects.

Two procedures are used to create an English word list. First, only voiceless aspirated stops /p', t', k'/ in the word-initial position are examined here due to the debatable implementation of English voiced stops; moreover, a CVCV sequence is used to ensure the target stop is stressed. Velar /k'/ followed by the high vowel /i/ is not included, as no corresponding words are found in Mandarin. Secondly, analogous to the Mandarin stimuli, disyllabic and not monosyllabic words are used to design the English word list.

3.3 Subjects

Thirty-six native speakers of Taiwanese Chinese were recruited from various departments at National Cheng Kung University in southern Taiwan. Subjects include 21 staff (mean age= 40 years) and fifteen students (mean age= 22 years), aged from 20 to 50 (mean age for all subjects= 32 years). All of the subjects were born and raised in Taiwan, have no marked regional accent, and reported no sophisticated knowledge of linguistics at the time of testing.

3.4 Procedures

Each subject was scheduled to record the word lists in a soundproof booth, using a high-quality microphone (AKG C1000S) and a professional 2-channel mobile digital recorder (MicroTrack 24/96). The target words for both languages were randomised in order not to be predictable. The recording was made when the subjects indicated they were ready. The subjects were first asked to read each word on the Mandarin and English word lists at a normal speed and repeat the whole lists twice in a row. All speakers were allowed to ask questions and practice words with which they were unfamiliar, but they were not informed of the purpose of the experiment. After the recording, they were asked to fill in a short questionnaire relating to their linguistic background.

3.5 Measurements and analyses

Wavesurfer software was used to make acoustic measurements of the speech material. Spectrograms and waveforms are displayed on screen and a manually controlled cursor is used for durational measurements, as shown in figure 1. VOT values were obtained by measuring the interval between the beginning of the release burst and the onset of the first formant visible in the frequency region. Target sounds that were obviously mispronounced are not included in the final analysis. Mean VOT values, standard deviations (SD), and graphical representation were made using EXCEL and SPSS. ANOVA tests were used for all statistical analyses, including the comparison of results and calculation of significance.

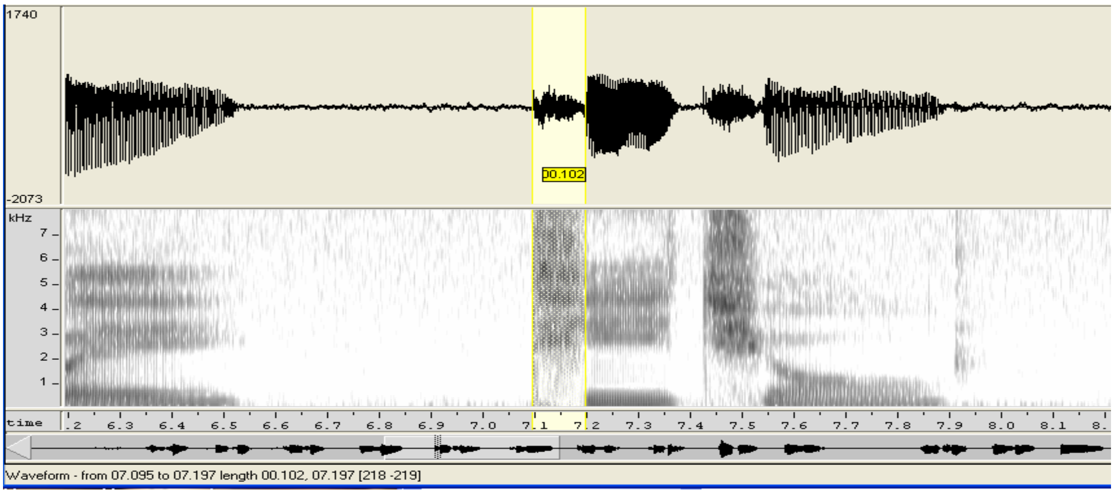

Figure 1: Spectrogram and waveform for the Mandarin word, “ti qiu”

4. Results

Mandarin VOT patterns for voiceless stops are discussed in section 4.1 below. Owing to the debatable phonetic implementations for English voiced stops, only voiceless aspirated stops (/p’, t’, k’/) in Mandarin and English are compared. Vowel quality is also taken into consideration in section 4.1.2, below.

4.1 Mandarin VOT

4.1.1 VOT means and distribution

The mean VOT values for six Mandarin stops are shown in figure 2, and detailed measurements including standard deviation (SD) are presented in table 3. Compared with the data reported by other researchers (Rochet & Fei, 1991; Liao, 2005; Chao et al., 2006), the VOT means for Mandarin stops presented in this study are relatively low, especially for the voiceless aspirated /k'/. Overall, VOT values for velar stops /k'/ and /k/ are significantly higher than those for bilabial and alveolar stops [$F (2, 835) = 15.917$, $p= .000< .05$]. Regarding the relation between place of articulation and VOT value, it is interesting to note that among voiceless aspirated stops, /t'/ has a higher value than /p'/, which does not conform to the general rule that VOT values rise as the place of articulation moves further back. The AONOVA test shows that the difference between /p'/ and /t'/ does not reach significance [$F (1,627) = 1.885$, $p= .170 > .05$]. However, this finding is only relevant to the voiceless aspirated /p', t'/, and not to the voiceless unaspirated /p, t/. In addition, as table 3 indicates, , contrary to English VOT patterns, the mean VOTs for Chinese bilabial and alveolar stops are much closer to each other, both for aspirated and unaspirated stops. The two main results of the present study are in accordance with studies by three other researchers (Rochet & Fei, 1991; Liao, 2005; Chao et al., 2006). The only difference is that for aspirated stops, Liao (2005) reports /p'/ with a significantly higher value than /t'/ [$F (1, 19) = 7.464$, $p= .013< .05$], while the two other studies showed no significant difference.

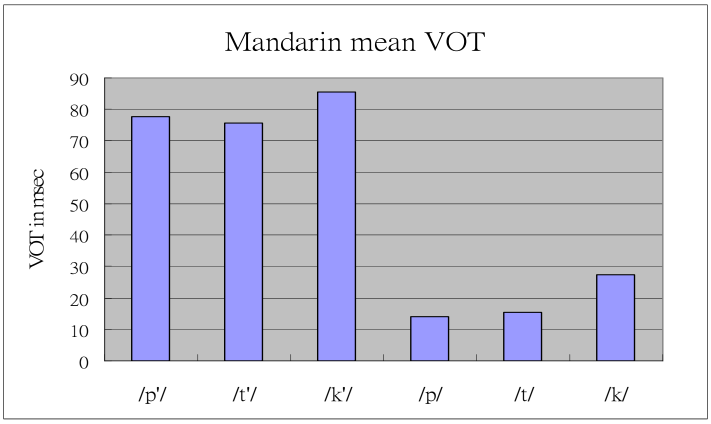

Figure 2. Mean VOT values for Mandarin stops

Table 3. General VOT means (ms) and standard deviation (SD) for all Mandarin stops

	/p'/	/t'/	/k'/	/p/	/t/	/k/
General means (in ms)	77.8	75.5	85.7	13.9	15.3	27.4
Standard deviation (SD)	23.7	18.4	19.4	6.6	5.7	9.6

Figures 3 and 4 show the VOT distribution for all Mandarin stops. Looking first at the voiceless aspirated stops, it can be seen that VOT ranges for /p', t', k/ are centralized around 63–90ms, 65–87ms, and 74–98ms, respectively. The values of standard deviation (SD) presented in table 3 also imply that /p'/ (SD=23.7 ms) allows more variation than /t/ (SD=18.4 ms) and /k'/ (SD=19.4 ms). As for voiceless unaspirated stops, the VOT ranges are centered around 10–18ms, 12–18ms, and 20–33ms, respectively. Unlike voiceless aspirated stops, the unaspirated /k/ (SD=9.6 ms) shows more variation than the two other stops and it may also be seen that the VOT range for /t/ is smaller than those for /p/ and /k/.

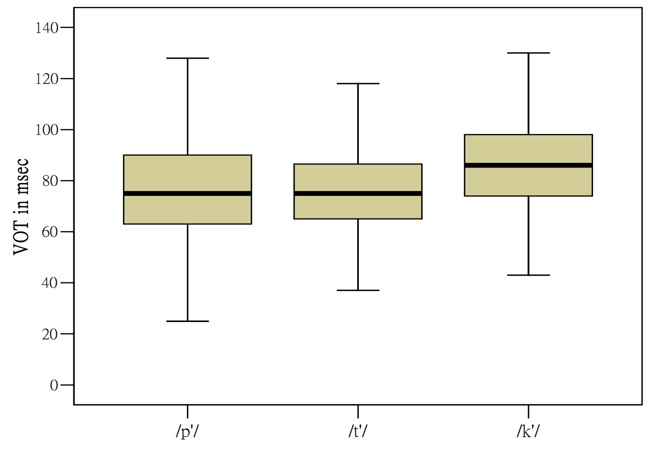

Figure 3. Boxplot for Mandarin voiceless aspirated stops

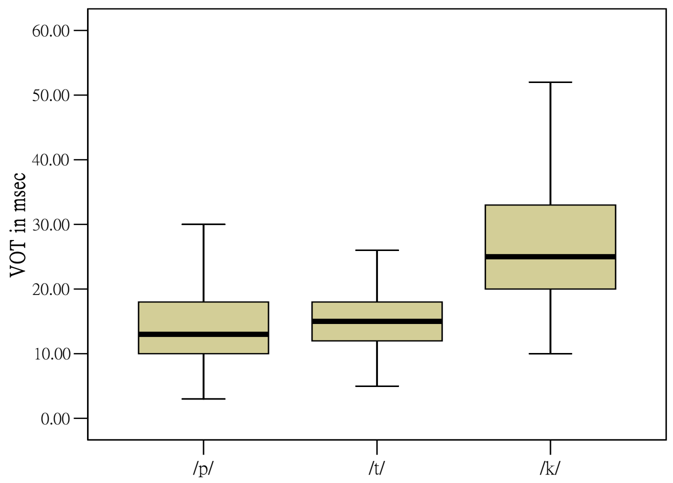

Figure 4. Boxplot for Mandarin voiceless unaspirated stops

4.1.2 Vowel context

Although there is an exception (Fant, 1973), it is widely accepted that word-initial stops have longer VOT values when followed by high vowels than by low vowels (Klatt, 1975; Weismer, 1979; Port, 1979; Rochet & Fei, 1991; Chao et al., 2006). In addition, Chao et al. (2006: 33) report that "all the stops, except /t/ which does not yield significance, have significantly longer VOTs when the following vowel is /i/ or /u/ than when it is /a/." Figures 5 and 6 show VOTs for voiceless stops followed by one of the three vowels, /i, u, a/. As the figures indicate, the VOTs for the unaspirated stops /p, t, k/ and the aspirated stops /p', t', k'/ are shorter when followed by the low vowel /a/ than by the high vowels /i/ and /u/. When doing t-test, the result also reveals that vowels, high or low, have significant effect on the VOTs for stops [p< .05].

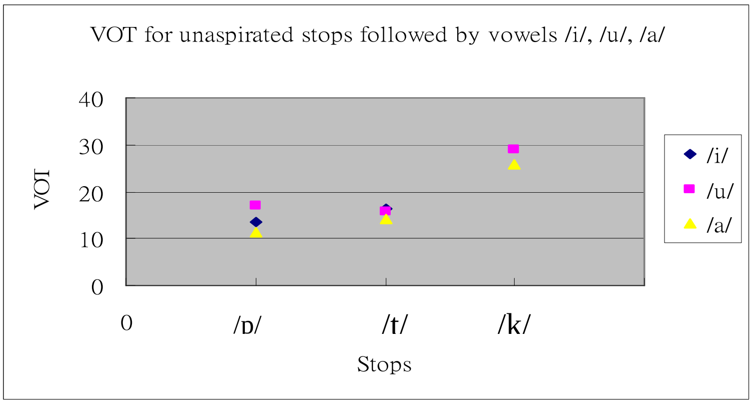

Figure 5: VOT for unaspirated stops followed by vowels /i/, /u/, /a/

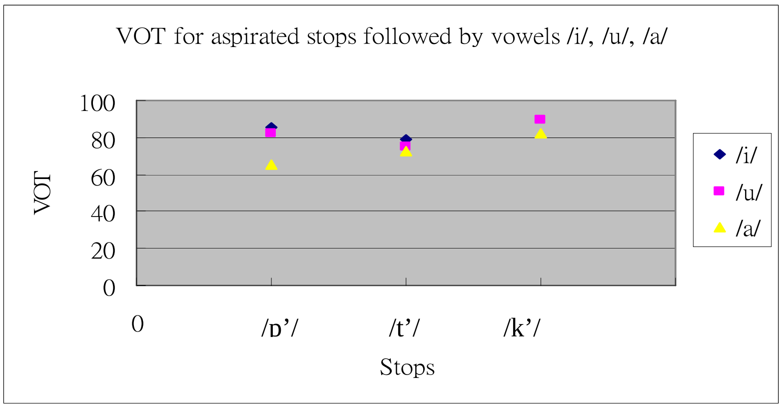

Figure 6: VOT for aspirated stops followed by vowels /i/, /u/, /a/

4.2 Comparing Mandarin and English VOT

As mentioned at the beginning of section 4, only phonetically voiceless aspirated stops are involved in the comparison of Mandarin and English VOT patterns. Figure 5 presents the mean VOTs for /p', t', k'/ in the two languages. The English mean VOTs are adopted from Lisker and Abramson's (1964) influential cross-language study. Visual inspection of the figure shows that Chinese speakers generally produce higher VOTs for /p', t', k'/ than English speakers. It should be noted that the differences between Mandarin and English VOTs are not stark but subtle, which raises the question whether L2 learners are aware of the slight differences between the two languages and are capable of producing them with authentic L2 production. This issue will be further discussed in section 5, below.

Apart from the differences mentioned above, place of articulation is another point which is worth noting. It is widely known that the further back the place of articulation, the longer the VOT, and there seems to be a general consensus on this. However, as figure 7 indicates, the mean VOTs for English /p', t', k'/ follow this rule, whereas Mandarin /p'/ and /t'/ do not. Moreover, the VOT values for aspirated bilabial and alveolar stops are closer to each other in Mandarin than in English VOT patterns.

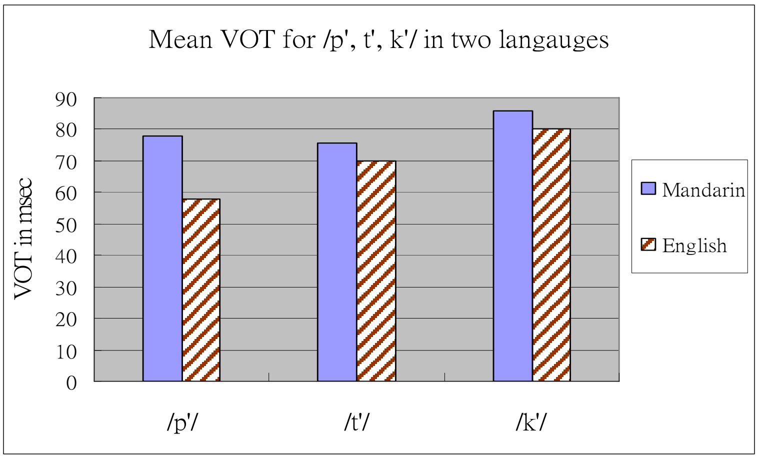

Figure 7. Mean VOTs for voiceless aspirated stops in Mandarin and English

4.3 English VOT in native Mandarin speakers

In section 4.2, it was mentioned that there are slight differences between VOT productions for voiceless aspirated stops in Mandarin and English. Since the two languages share similar VOT patterns with only subtle differences, it is worth investigating how native Chinese speakers produce English voiceless aspirated /p', t', k'/. Chao and Chen (2006) find that native Chinese speakers often produce English /p', t', k'/ with 'compromise' values. Thus, it is interesting to observe the English VOT patterns of the L2 learners (i.e. Chinese learners of English) in this study.

4.3.1 VOT means and distribution

Figure 8 shows the mean VOT durations for English /p', t', k'/ produced by native Chinese speakers; detailed measurements including SD are presented in table 4. As the figure shows, the velar /k'/ has a highly significantly longer VOT than the three voiceless aspirated stops [$F_{(2,831)} = 106.450$, $p = .000 < .05$]. Nevertheless, VOT values for /p'/ and /t'/ still do not reach significance ($p > .05$). This result is similar to that found for Mandarin VOT patterns that differ in place of articulation. As mentioned above, the VOT values for Mandarin /p', t'/ do not increase as the place of articulation moves further back. However, Chinese speakers' L2 production accords with the general rule.

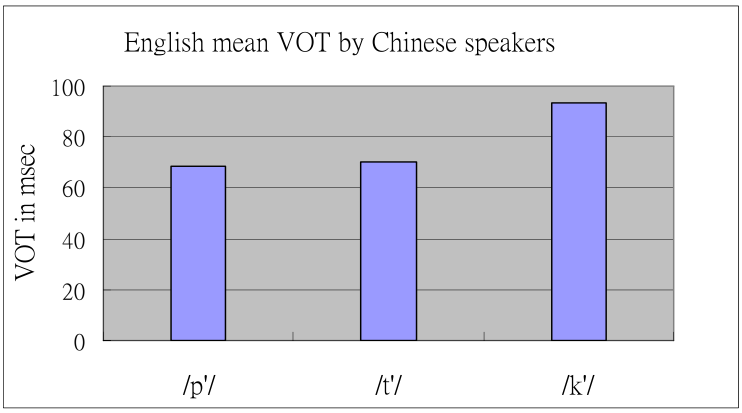

Figure 8. English VOT means for voiceless aspirated stops by Chinese speakers

Table 4. English VOT means (ms) produced by native Chinese speakers; standard deviation (SD) for voiceless aspirated stops

	/p'/	/t'/	/k'/
General means (in ms)	68.7	70.2	93.4
Standard deviation (SD)	21.8	19.2	20.5

Figure 9 compares the mean VOTs for English productions by native Chinese speakers, native Mandarin productions, and native English productions. Looking at the figure, it may be noted that native Chinese speakers produce intermediate VOT values only for English aspirated /p'/, by comparison with native speaker productions for either language. One may also notice that in their production of aspirated velar /k'/, Chinese speakers produce far higher English VOTs than in their corresponding Chinese production and than the English mean produced by native speakers. As for /t'/ production, it is interesting to observe that English mean VOTs for native speakers and Chinese subjects are almost the same (VOT= 70ms for the former; VOT= 70.2ms for the latter). Individual variations among Chinese native speakers should also be taken into account when forming comparisons.

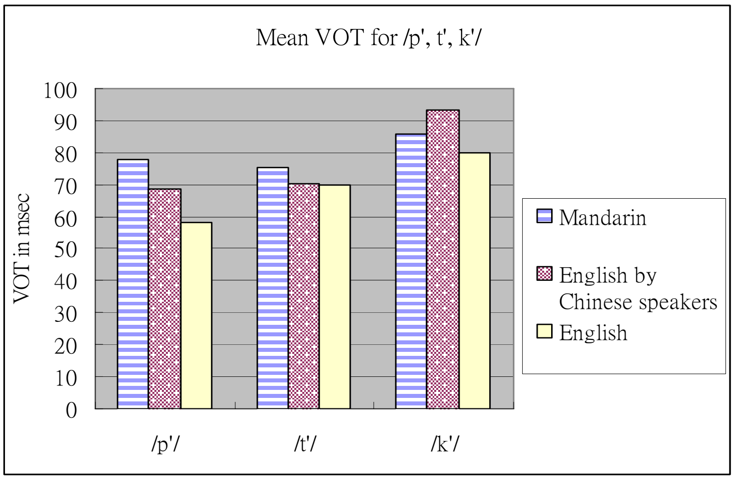

Figure 9. Mean VOTs for voiceless aspirated stops in Mandarin, English produced by Chinese speakers, and English produced by native speakers

5. Discussion

Three important conclusions may be derived from the present study and will be discussed in detail below.

With respect to Mandarin VOT patterns, the VOT means obtained for the six Chinese stops are somewhat lower than the data reported in previous studies (Rochet & Fei, 1991; Liao, 2005; Chao et al., 2006), as shown in table 5, below. The lowness of these values may be explained as follows. First, disyllabic words were used in the present study, rather than the monosyllables which were examined in the study by Rochet & Fei (1991). Using disyllables creates a more natural context for the subjects and likely obtains more accurate VOT values. Methodological differences may be another reason for the lower values. A third explanation is the number of subjects tested: the experiment for the present study used more subjects than the previous two studies (Liao, 2005; Chao et al., 2006), which means the values obtained are probably more reliable.

As for the comparison of English and Mandarin VOT patterns, the results indicate that for voiceless aspirated stops, both English and Mandarin belong to the long-lag category, contrary to previous findings. Chao et al. (2006) claim that Mandarin /p', t', k'/ fall into the highly aspirated category and suggest that the aspirated category should not be considered as a single long continuum. The VOT means reported by Rochet and Fei (1991) also imply that Mandarin and English occupy different regions of the VOT continuum. Although both languages share similar stop category, there are still differences between them. Comparing /p'/ productions first, it can be seen that Chinese speakers produce much longer VOT values than native English speakers. Among the three voiceless aspirated stops in each language, only the alveolar /t'/ has a value close to the others. This accords with the results presented by other researchers who examined Chinese voiceless stops (Rochet & Fei, 1991; Liao, 2005; Chao et al., 2006). Place of articulation is another factor worth noting. Although the general rule states that the further back the place of articulation, the longer the VOT, this is not the case for Mandarin voiceless aspirated stops. The results indicate that the VOT for the aspirated alveolar stop /t'/ is shorter than that for the aspirated labial stop /p'/, except when they are followed by the low vowel /a/. The results for stops followed by the low vowel /a/ are consonant with the results of Chen, Tsay, and Hong's study (1998). The cause of this result is complicated and requires further discussion. In addition to Mandarin, Lisker and Abramson (1964) report that unaspirated stops in Tamil and aspirated stops in Cantonese and Eastern Armenian do not follow the general rule either. Of these four languages, both Cantonese and Mandarin are tone languages. Whether tone affects VOT values is still a controversial question. Some researchers have claimed that there is no significant influence (Chen et al. 1998; Ran, 2005), whereas in a study by Liu et al. (*Article in Press*) it is found that "VOT values associated with high-level and high-falling tones were shorter than those associated with mid-rising and falling-rising tones." The test stimuli used in the present study are not in the same tone; therefore, if tones do influence VOT values, it is possible that some of the results may be explained in this way.

Table 5. Mean VOT values (ms) for Mandarin voiceless stops

	Rochet & Fei, 1991 (monosyllables)	Liao, 2005 (disyllables)	Chao et al., 2006 (disyllables)	Present study (disyllables)
	Mean	**Mean**	**Mean**	**Mean**
/p'/	99.6	75.4	82	77.8
/t'/	98.7	71.4	81	75.5
/k'/	110.3	98.8	92	85.7
/p/		17.9	14	13.9
/t/		18.6	16	15.3
/k/		28	27	27.4

As for vowel context, it is found that the VOTs for stops, both unaspirated and unaspirated, are longer when followed by the high vowels /i/ and /u/ than by the low vowel /a/. This supports the findings of many studies (Port, 1979; Gósy, 2001; Rochet & Fei, 1991; Chao et al., 2006). Although there are some exceptions (Lisker & Abramson, 1967; Fant, 1973), more and more studies support the view that high/low vowel quality influences the VOT value of preceding stops. Front/back vowel quality has no significant influence on VOT.

Since the differences between Mandarin and English VOTs are subtle, it is worth observing the English VOT performance of Chinese speakers. Chao and Chen (2006) propose that native Chinese speakers often produce English /p', t', k'/ with 'compromise' VOT values. Whether these speakers are able clearly to distinguish the subtle differences between the two languages, or whether their L2 productions are influenced by their first language (i.e. Mandarin), is an interesting issue for further discussion. The present findings reveal that, except for /k'/, Chinese speakers' L2 productions of /p'/ and /t'/ are either intermediate or close to English native speakers' productions. To understand the exception of /k'/ values, language proficiency could be taken into consideration. Liao (2005) suggests that proficiency has a certain influence on interlanguage production of stop consonants. According to Liao (2005), L2 learners with a higher level of proficiency have greater accuracy than those with a lower level. 21 of the staff members observed in this study are classified as having a low level of proficiency, which may be one of the reasons for their striking /k'/ production. It should also be noticed that the mean VOT values of English /p'/ and /t'/ by Chinese speakers are close to each other. Previous studies have examined this phenomenon and provided various suggestions for factors affecting L2 production. On the one hand, it is suggested that first language (L1) effect on L2 plays a crucial part in L2 learners' VOT productions (Thompson, 1991; Flege et al., 1997). On the other hand, Flege and Hammond (1982) also claim that speakers actually produce intermediate phonetic categories between their native language and a foreign language. Variations in L2 production both for the same persons over time and for different speakers could be examined further and taken into consideration in future studies.

References

[1] A. Caramazza, G. Yeni-Komshian, E. Zurif, and E. Carbone, "The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals," *Journal of the Acoustical Society of America*, vol. 54, pp. 421–28, 1973.

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

[3] D. H. Klatt, "Voice Onset Time, frication, and aspiration in word-initial consonant clusters," *Journal of Speech and Hearing Research*, vol. 18, pp. 686–706, 1975.

[4] D. Kewley-Port, D. B. Pisoni, and M. Studdert-Kennedy, "Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants," *Journal of Acoustical Society of America*, vol. 73, no. 5, pp. 1779–1793, 1983.

[5] G. H. Yeni-Komshian, A. Caramazza, and M. S. Preston, "A study of voicing in Lebanese Arabic," *Journal of Phonetics*, vol. 7, pp. 197–204, 1977.

[6] G. Khattab, "VOT production in English and Arabic bilingual and monolingual children," *Leeds working papers in linguistics and phonetics*, vol. 8, pp. 95–122, 2000.

[7] G. Weismer, "Sensitivity of voice-onset-time (VOT) measures to certain segmental

features in speech production," *Journal of Phonetics*, vol. 7, pp. 197–204, 1979.
[8] G. J. Docherty, *The timing of voicing in British English obstruents.* New York: Foris, 1992.
[9] 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," *Article in Press.*
[10] I. Thompson, "Foreign accents revisited: the English pronunciation of Russian immigrants," *Language Learning*, vol. 41, pp. 177–204, 1991.
[11] J. E. Flege and R. M. Hammond, "Mimicry of non-distinctive phonetic differences between language varieties," *Studies in Second Language Acquisition*, vol. 5, no. 1, pp. 1–17, 1982.
[12] J. E. Flege, O.-S. Bohn, and S. Jang, "Effects of experience on non-native speaker's production and perception of English vowels," *Journal of Phonetics*, vol. 25, pp. 437–470, 1997.
[13] J. R. Benkí, "Place of articulation and first formant transition pattern both affect perception of voicing in English," *Journal of Phonetics*, vol. 29, pp. 1–22. 2001.
[14] K. Chen, J. Tsay, and G. Hong, "Duration of initials in Mandarin: Fundamental acoustic research and its clinical significance," *Journal of Speech Language-hearing Association*, vol. 13, pp. 138–149, 1998.
[15] K.- Y. Chao, G. Khattab, and L.- M. Chen, "Comparison of VOT Patterns in Mandarin Chinese and in English," in *Proceedings of the 4th Annual Hawaii International Conference on Arts and Humanities*, 2006, pp. 840–859.
[16] L. Lisker and A. S. Abramson, "A cross-language study of voicing in initial stops: acoustical measurements," *Word*, vol. 20, pp. 384–422, 1964.
[17] L. Lisker and A. S. Abramson, "Some effects of context on voice onset time in English stops," *Language Speech*, vol. 10, pp. 1–28, 1967.
[18] M. Gósy, "The VOT of the Hungarian voiceless plosives in words and in spontaneous Speech," *International Journal of Speech Technology*, vol. 4, pp. 75–85, 2001.
[19] M. M. Kehoe, C. Lleó, and M. Rakow, "Voice onset time in bilingual German-Spanish children," *Bilingualism: Language and Cognition*, vol. 7, pp. 71–88, 2004.
[20] O.-S. Bohn and J. E. Flege, "Perceptual switching in Spanish/English bilinguals," *Journal of Phonetics*, vol.21, no. 3, pp. 267–290, 1993.
[21] P. A. Keating, "Phonetic and phonological representation of stop consonant voicing," *Language*, vol. 60, pp. 286–319, 1984.
[22] P. A. Keating, W. Linker, and M. Huffman, "Patterns in allophone distribution for voiced and voiceless stops," *Journal of Phonetics*, vol. 11, pp. 277–90, 1983.
[23] Q. B. Ran, "Experimental studies on Chinese obstruent consonants: with the emphasis on standard Chinese," Ph. D. dissertation, Nankai University, Mainland China, 2005.
[24] R. F. Port and R. Rotunno, "Relation between Voice-Onset Time and vowel duration," *Journal of the Acoustical Society of America*, vol. 66, pp. 654–62, 1979.
[25] R. H. Kessinger and S. E. Blumstein, "Effects of speaking rate on voice-onset time in Thai, French, and English," *Journal of Phonetics*, vol. *25*, pp. 143–168, 1997.
[26] S. J. Liao, "Interlanguage production of English stop consonants: A VOT analysis," M. A. thesis, National Kaohsiung Normal University, Kaohsiung, Taiwan, 2005.
[27] T. Cho and P. Ladefoged, "Variation and universals in VOT: evidence from 18 languages," *Journal of Phonetics*, vol. 27, pp. 207–29, 1999.
[28] T. J. Riney and N. Takagi, "Global foreign accent and voice onset time among Japanese EFL speakers," *Language Learning*, vol. 49, no. 2, pp. 275–302, 1999.
[29] T. J. Riney, N. Takagi, K. Ota, and Y. Uchida, "The intermediate degree of VOT in Japanese initial voiceless stops," *Journal of Phonetics*, vol. 35, no. 3, pp. 439–443, 2007.
[30] X.-R. Zheng and Y.-H. Li, "A contrastive study of VOT of English and Korean Stops," *Journal of Yanbian University*, vol. 38, no. 4, pp. 99–102, 2005.