

# A Cross-Linguistic Study of Voice Onset Time in Stop Consonant Productions

Kuan-Yi Chao\* and Li-mei Chen\*

## Abstract

This study examines voice onset time (VOT) for phonetically voiceless word-initial stops in Mandarin Chinese and in English, as spoken by 11 Mandarin speakers and 4 British English speakers. The purpose of this paper is to compare Mandarin and English VOT patterns and to categorize their stop realizations along the VOT continuum. As expected, the findings reveal that voiceless aspirated stops in Mandarin and in English occur at different places along the VOT continuum and the differences reach significance. The results also suggest that the three universal VOT categories (*i.e.* long lead, short lag, and long lag) are not fine enough to distinguish the voiceless stops of these two languages.

**Keywords:** Voice Onset Time (VOT), Voiceless Stops

## 1. Introduction

Over the past few decades, beginning with Lisker and Abramson's [Lisker and Abramson 1964] study, a considerable number of studies have investigated voicing contrasts in stops by the use of voice onset time (VOT). VOT has come to be regarded as one of the most important methods for examining the timing of voicing in stops (especially in word-initial position) and has been applied in studies of many languages. However, only a few attempts have been made to examine VOT patterns in Mandarin so far. Three universal categories of phonetically voiceless stops are generally recognized [Lisker and Abramson 1964], and Mandarin and English occupy the same place along the VOT continuum according to this general categorization. Nevertheless, differences between voiceless stops in Mandarin and English do exist. Since no existing studies compare the VOT patterns of these two languages, the aim of the present study is to provide a comparison of phonetically voiceless stops in Mandarin and in English, and to pinpoint the differences between their VOT patterns.

---

\*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.

## 2. Literature Review

Voicing contrast in stops has been widely discussed in phonetics and phonology. Voice Onset Time (VOT), the acoustic cue used to measure the timing of voicing, was first described by Lisker and Abramson in their well-known cross-language study of voicing in initial stops in 11 languages. According to Lisker and Abramson, voice onset time serves as a device for 'separating the stop categories of a number of languages in which both the number and phonetic characteristics of such categories are said to differ.' The authors also indicate that the measure of VOT is found to be highly effective in separating phonemic categories, such as voiced and voiceless, although the languages under study differ both in the number of those categories and their phonetic features. For example, in English, the minimal pair 'pan' and 'ban' can only be distinguished by voicing contrast. Thus, it can be seen that VOT plays an important role in differentiating voiced from voiceless stops, especially for lexical purposes.

### 2.1 VOT Definition

Lisker and Abramson's study defined VOT as 'the time interval between the burst that marks release of the stop closure and the onset of quasi-periodicity that reflects laryngeal vibration' ([Lisker and Abramson 1964: 422]), and used the concept to examine word-initial stops in 11 languages. Since then, a considerable number of studies of many languages has been undertaken, including a report on VOT in 51 languages [Keating *et al.* 1983], and another more recent study on VOT in 18 languages [Cho and Ladefoged 1999].

Although VOT is now in widespread use for measuring the timing of voicing in stops, its role as a reliable measure to distinguish between voiced and voiceless stops has been brought into question. Bohn and Flege's [Bohn and Flege 1993] findings suggest that VOT, the acoustic parameter of voicing contrasts in word-initial stops, may not be as important to the perception of stop voicing as is commonly supposed. Docherty [Docherty 1992] argues that voice onset time focuses narrowly on the timing of voicing in word-initial stops and does not take into account stops in word-final and word-medial positions. Caramazza, Yeni-Komshian, Zurif, and Carbone [Caramazza *et al.* 1973] also conclude that VOT is an 'insufficient' cue to the voicing contrast for French-English bilinguals.

On the other hand, some researchers argue that other acoustic cues play a role. For example, Klatt [Klatt 1975: 695] suggests that there are five equally important acoustic cues in English other than voice onset time, namely, low frequency energy in following vowels, burst loudness, fundamental frequency, pre-voicing, and segmental duration. Nevertheless, despite some research showing the limitations of voice onset time, it is still regarded as one of the most important acoustic parameters and has been used extensively in measuring word-initial stops.

## 2.2 VOT Category

Lisker and Abramson [Lisker and Abramson 1964: 388] examine 11 languages and classify them into three groups according to the number of stop categories each language contains. They also suggest that each stop category falls into one of three ranges,  $-125$  to  $-75$  ms,  $0$  to  $+25$  ms, and  $+60$  to  $+100$  ms, respectively (p. 403). Following Lisker and Abramson's categorization, both Mandarin and English fall into the two-category group of languages and occupy the same range along the VOT continuum, that is,  $0$  to  $+25$  ms for [p, t, k] and  $+60$  to  $+100$  ms for [p', t', k'].

However, this classification is too general to note the subtle variations between the two languages. Cho and Ladefoged [Cho and Ladefoged 1999: 223] go into further detail and classify the range for voiceless aspirated and unaspirated occlusives, concentrating particularly on velar stops across 18 languages. They distinguish four categories, which they name *unaspirated* (velar stops with a mean VOT of around 30 ms), *slightly aspirated* (with a mean VOT of around 50 ms), *aspirated* (with a mean VOT of around 90 ms), and *highly aspirated* (with a mean VOT of over 90 ms). According to Cho and Ladefoged [Cho and Ladefoged 1999: 223], Mandarin and English do not occupy the same place along the VOT continuum, especially for voiceless aspirated stops. Further discussion of VOT categories for Mandarin and English stops is in Section 2.4, below.

## 2.3 Effect on VOT

Voice onset time (VOT) is known to vary with place of articulation. The general principle is that, the further back the place of articulation, the higher the VOT values [Fischer-Jorgensen 1954; Lisker and Abramson 1964; Docherty 1992; Cho and Ladefoged 1999]. Cho and Ladefoged suggest several ways of explaining this principle, including explanations based on the laws of aerodynamics, articulator movement and differences in mass of articulators. According to the explanation based on the laws of aerodynamics, the main reason for velar stops having a longer VOT than alveolar or bilabial stops is the relative size of the supraglottal cavity behind the constriction. With the velar stop, greater air pressure builds up in the vocal tract because the supraglottal cavity becomes smaller and it takes longer for the pressure to fall at the beginning of the release phase.

In accordance with Lisker and Abramson's findings, velar stops have consistently higher VOT values than the other stops. Rochet and Fei's [Rochet and Fei 1991] study, which examines Mandarin stops, shows the same trend — velar stops consistently show the longest VOT. They also find that, in Mandarin voiceless aspirated stops (*i.e.* [p', t', k']), the apical stop is correlated with slightly lower values than the labial one; this does not conform to the general agreement.

Vowel quality is another intrinsic effect which plays a crucial part in affecting voice

onset time. Although Lisker and Abramson claim that vocalic environment does not have a major influence on VOT, a number of reports have questioned their claim [Klatt 1975; Weismer 1979; Port and Rotunno, 1979]. Generally, it has been found that tense high vowels have longer VOTs than lax low vowels [Klatt 1975; Weismer 1979; Port and Rotunno 1979]. However, owing to language-specific variation, the correlation between voice onset time and vowel quality does not allow any definite conclusions.

Rochet and Fei state briefly that the mean VOT for both sets of Mandarin stops have greater values when they are followed by a high vowel /i/ or /u/ rather than the low vowel /a/. The study provides information on the phonetic features of VOICED and VOICELESS stops in Mandarin. (In this study, the upper-case forms — ‘VOICED’ and ‘VOICELESS’ — will be used to refer to stops’ phonological status, while the lower-case forms — ‘voiced’ and ‘voiceless’ — refer to their phonetic type.) More instrumental studies are needed in order to establish more complete and reliable Mandarin VOT patterns.

## 2.4 Mandarin and English Stops and VOT Patterns

As mentioned above, a sizable body of studies has been carried out to investigate the phonetic characteristics of voiced and voiceless stops in various languages using voice onset time as an important acoustic cue. Most existing studies concentrate on English [Lisker and Abramson 1964; Klatt 1975; Port and Rotunno 1979; Weismer 1979; Keating *et al.* 1983; Docherty 1992]. Other languages examined include Spanish [Lisker and Abramson 1964; Flege and Hammond 1982; Flege and Eefting 1987; Fellbaum 1996], French [Caramazza *et al.* 1973; Rochet *et al.* 1987], Arabic [Flege 1980; Khattab 2000] and Japanese [Shimizu 1990; Riney and Takagi 1999]. Among these investigations, there is very little data available on VOT for Mandarin word-initial stops. Thus, it does not appear that Mandarin VOT patterns have been examined extensively; to our knowledge, only Rochet and Fei have examined Mandarin Chinese.

This study will only discuss syllable initial stops, owing to the absence of stops in any other position in Mandarin Chinese. It is known that, in word-initial position, English VOICED stops are voiced or voiceless and unaspirated, and that VOICELESS stops are voiceless and aspirated [Keating *et al.* 1983; Keating 1984; Docherty 1992]. Although there are two possible phonetic implementations of English VOICED stops, Keating [Keating 1984: 43] indicates that ‘English divides up the VOT continuum with some lead values but mainly short lag vs. long lag.’ In Lisker and Abramson, VOT measurements occurring before the release burst are assigned negative values and called *voicing lead*, while VOT measurements occurring after the release burst are assigned positive values and called *voicing lag*. Lisker and Abramson [Lisker and Abramson 1964: 395] also provide two sets of values for English voiced stops (VOTs with lead and with short lag) and suggest that only a single type is

produced by each native speaker. Based on the distinction of Keating and Lisker and Abramson, English is described as having, in general, short lag and long lag VOT patterns.

In comparison with English, Mandarin shows less variation in implementation. All Mandarin stops are phonetically voiceless and are only differentiated by aspiration. According to data provided by Rochet and Fei, VOT duration for Mandarin [p', t', k'] ranges between 90 and 110 ms, while that of Mandarin [p, t, k] ranges between 10 and 25 ms (depending on the place of articulation). In Keating's study, Mandarin and English stops are classified as phonetically the same, and both fall into short lag vs. long lag patterns; however, there remain some subtle differences between the two languages.

Table 1 shows detailed measurements of English VOT means and ranges, including American English [adopted from Lisker and Abramson 1964] and British English [adopted from Docherty 1992]. The VOT pattern in Mandarin Chinese [adopted from Rochet and Fei 1991] is presented in Table 2. According to Cho and Ladefoged's definition of voiceless unaspirated stops, English [p, t, k] fall into the 'unaspirated' range with a VOT of under 30 ms. However, for voiceless aspirated stops (*i.e.* [p', t', k']) Mandarin occupies a 'highly aspirated' position along the VOT continuum, while English lies in the 'aspirated' region. The table also indicates that although [p', t', k'] in Mandarin and English fall into two categories, they are not completely different because the VOT ranges in these two languages overlap.

**Table 1. VOT means and ranges in English.**

	Lisker and Abramson 1964 (AE)		Docherty 1992 (BE)	
	Mean	Range	Mean	Range
<b>p'</b>	58	20 -120	42	10 - 80
<b>t'</b>	70	30 -105	64	30 - 110
<b>k'</b>	80	50 -135	62	30 - 150
<b>p</b>	1	0 - 5	15	0 - 50
<b>t</b>	5	0 - 25	21	0 - 50
<b>k</b>	21	0 - 35	27	10 - 60

(AE=American English; BE=British English; all measurements are in milliseconds (ms). Note: In this study, based on Keating [Keating 1984] and Lisker and Abramson's [Lisker and Abramson 1964] distinction, [p', t', k'] is used for referring to voiceless aspirated stops while [p, t, k] is used as voiceless unaspirated stops in phonetic distinctions.)

**Table 2. VOT means for voiceless aspirated stops in Mandarin.**

Rochet and Fei 1991	
Mean ( in ms)	
<b>p'</b>	99.6
<b>t'</b>	98.7
<b>k'</b>	110.3

(Note: Rochet and Fei provide the mean VOT for voiceless aspirated [p', t', k'] but not for voiceless unaspirated [p, t, k], nor do they provide VOT ranges.)

### 3. Methodology

#### 3.1 Aims of the Experiment

As mentioned above, a few instrumental studies have examined VOT in Mandarin. No existing studies even attempt to compare VOT patterns in Mandarin and English. Therefore, the present experiment examines subtle differences in VOT production between the two languages and seeks to determine whether Mandarin and English stop realizations occupy the same place along the VOT continuum.

#### 3.2 Linguistic Material

Stops in Mandarin Chinese occur only in the initial position. The present experiment examines two sets of phonetically voiceless stops, that is, aspirated [p', t', k'] and unaspirated [p, t, k]. Each of the stops is followed by three peripheral vowels in turn: two high vowels, /i/ and /u/, and one low vowel, /a/, thus giving a total of three variations for each stop. Two exceptions to this are the velar stops [k'] and [k], as no Mandarin lexical items exist for the sequences [k'] or [k] followed by the high front vowel /i/. All the words used are real words.

Unlike English, every character in Mandarin is correlated with monosyllabic sounds only, all of which can exist independently. However, two or more characters usually stand side by side to form a 'word' which is more complete and more meaningful. Take 'da ying' for example, the sound 'da' can have various meanings; however, 'da ying' means 'to promise' and is easy for experiment participants to understand. Thus, in the present experiment, compound words are used rather than single characters (as in previous experiments, *e.g.*, Rochet and Fei 1991) because they make more sense to the subjects, and because compound words fit the testing format better than single characters. The inventory of all stimuli words is listed in Appendix A.

The following procedure was followed to create English and Mandarin word lists. First, the word-lists were oriented to the Mandarin Chinese lexicon in order to obtain comparable material in both English and Mandarin. Thus, all the English stops examined here are in the word-initial position and the target words contain two high vowels (/i/ and /u/) and one low vowel (/a/). Velar stops [k'] and [k] followed by the high front vowel /i/ are not included as they do not occur in Mandarin. Secondly, disyllabic words were chosen rather than monosyllabic materials because the Mandarin word-list consists of pairs of characters put together to make target words. Every effort was made to find Mandarin and English words in which the stops were followed by 'similar' vocalic and consonantal contexts; however, it was very difficult to create a list of homophones across the two languages owing to phonetic and phonological differences.

### **3.3 Participants**

11 Mandarin speakers participated in this experiment: all were females aged between 19 and 35 years (mean=27). All of the participants were born in Taiwan, where Mandarin Chinese is the predominant language. None of them had a marked regional accent, although they were raised in different areas of Taiwan. All the Mandarin participants were college students. Four female native speakers of British English participated in this experiment as a control group. The native English speakers were older (mean=32) than the native speakers of Mandarin. All were born and raised in the UK. The Mandarin speakers participated in the Mandarin-based test, while the English speakers participated in the English-based test.

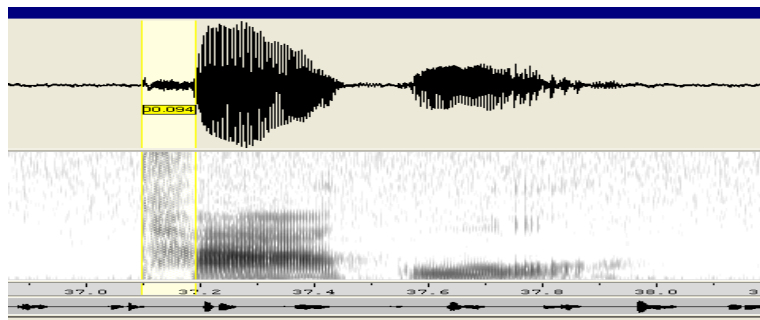
### **3.4 Procedures**

In order to ensure the target sounds were not predictable for any participants who might have linguistics backgrounds, all the speech materials, both Mandarin and English, were randomized. Target words were equally divided into two subgroups in each language due to recording limitations. Two preparation words were added to each group, as the first and last items, in order to allow participants to practice before the target words being recorded. One by one, participants were recorded in a sound booth, and their readings recorded straight into a computer. The participants were asked to start reading each word when they were ready; the list of words was read five times in a row at a comfortable speed. The first iteration was not used for analysis. This was to ensure that participants were familiar with the entire recording procedure before they produced a recording which would actually be analyzed. All speakers were given instructions in English, and Mandarin participants also received instructions in Mandarin. The participants were told that the object of the experiment was to examine speech, but they were not told that their production of specific sounds (*i.e.* stops) would be assessed.

### **3.5 Measurements and Analyses**

Acoustic measurements of the speech material were made using Wavesurfer software. Following Lisker and Abramson [Lisker and Abramson 1964: 422], VOT was measured as the interval between the beginning of the release burst and the onset of quasi-periodicity that reflects glottal vibration in F1 in the following vowel. Spectrographic measurements were taken of the VOTs of Mandarin initial stops in a total number of 800 tokens. Mean VOT values were calculated for the stops produced by each participant. A few measurements were missing owing to discarded tokens or imperceptible release bursts which made it difficult to measure VOT. The analysis involved displaying two panels (spectrogram and waveform) on separate portions of the screen, and using a manually controlled cursor for durational measurements, as shown in Figure 1. Measurement reliability was assessed by re-measuring three randomly selected stops in each group from duplicate spectrograms. The average

difference between the two measurements was 2.2 ms, with a range of 0-5 ms. T-tests were used for comparison of results and calculation of statistical significance. In this study, the p values which were less than 0.05 were reported to be significant.



*Figure 1. Example of a test word “ta guo” (踏過) shown in the acoustic waveform (top panel) and spectrogram (bottom panel).*

## 4. Results

### 4.1 Overall Results for Mandarin VOT

#### 4.1.1 VOT Means and Distribution

Table 3 presents the mean VOT values and ranges for each of the six Mandarin stops. It was found that the velar stops [k'] and [k] have higher VOT values than the other stops. This supports the findings in Lisker and Abramson [Lisker and Abramson 1964] in their cross-language study. The mean VOT value for [t'] is slightly lower than that for [p'], which does not conform to the general agreement that the further back the place of articulation, the longer the VOT. However, the difference between [p'] and [t'] does not reach significance. This concurs with Rochet and Fei's finding [Rochet and Fei 1991] that the mean VOTs for [p'] and [t'] do not differ significantly from each other. T-tests reveal that the VOT value for the velar stop [k'] is significantly higher than those for [p'] and [t'].

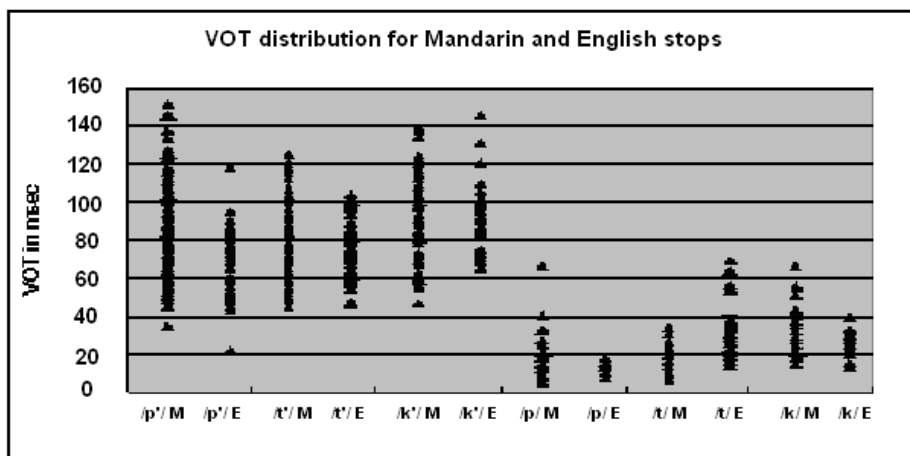
Regarding the voiceless unaspirated stops, it is noticeable that the mean VOT for [k] is higher than that for [t], which in turn is higher than the VOT for [p]. Compared with voiceless aspirated stops, unaspirated [p, t, k] are in closer accord with the consensus reported in the literature. T-tests were carried out to determine if the differences between [p], [t], and [k] were statistically significant. The results show that the VOT for the velar stop is significantly higher than those of the other two stops, and there is a significant difference between [p] and [t] as well. Overall, the present study's findings on Mandarin VOT patterns are in accordance with the study by Rochet and Fei, in which the authors found that the mean VOT for [t'] was slightly but not significantly lower than that for [p'].



**Table 3. VOT ranges and general means (in ms) for Mandarin stops.**

	p'	t'	k'	p	t	k
Min	35	45	50	7	7	15
Max	147	123	138	65	33	65
General means	82	81	92	14	16	27

VOT distribution for all stops in Mandarin and English is shown in Figure 2. As the figure indicates, the VOT ranges for voiceless aspirated [p', t', k'] are 35-147 ms, 45-123 ms and 50-138 ms, respectively. However, it is clear that the VOT distribution for [p'] tends to center in the range from 45 to 130 ms, and that for [k'] is concentrated in the range from 58 to 125 ms. As for voiceless unaspirated stops, the VOT ranges for [p, t, k] are 7-65 ms, 7-33 ms and 15-65 ms, respectively. VOT distribution is centered in the range of 7-28 ms for [p] and 15-50 ms for [k]. It is interesting to note that the ranges for [t'] and [t] are narrower than those for the other stops; that is, the bilabials and the velars allow more variation than [t'] and [t].



**Figure 2. VOT distribution for all stops in Mandarin (M) and English (E).**

### 4.1.2 Vowel Context

As mentioned above, vowel quality is another important factor influencing voice onset time. It is now widely accepted that tense high vowels are correlated with longer VOTs than lax low vowels [Klatt 1975; Weismer 1979; Port and Rotunno 1979]. Table 4 shows the mean VOTs for Mandarin stops followed by three peripheral vowels /i, a, u/. It should be noted that there are no lexical items with the sound sequences of /ki/ or /k'i/ in Mandarin. As shown in the table, all the stops have longer VOTs when they are followed by a high vowel, /i/ or /u/, than by the low vowel /a/. A series of t-tests were carried out to determine if the differences are statistically significant in the VOT values between the stops when followed by different vowels. The results indicate that all the stops, except [t'] which does not reach significance,

have significantly longer VOTs when the following vowel is /i/ or /u/ than when it is /a/. The present results are in keeping with the correlation between VOT and vowel quality reported in Rochet and Fei's findings. To sum up, it may safely be assumed that Mandarin stops generally fit in with the assumption that tense high vowels contribute to longer VOTs in preceding stops than lax low vowels.

**Table 4. Mean VOT values (in ms) for Mandarin stops with three following vowels /i/, /a/ and /u/**

	p'	t'	k'	p	t	k
i	90	82	X	13	18	X
u	87	82	98	18	17	33
a	70	81	86	12	14	22

#### 4.2 Overall Results for English VOT

Table 5 shows the mean VOT values produced by the British English speakers. English mean VOTs with and without subject A are juxtaposed for comparison. As can be clearly seen, the place of articulation has an effect on VOT for [p', t', k'] but not for [p, t, k]. Regarding the mean VOTs by native English speakers with subject A, it is found that the mean VOT for aspirated velar [k'] is significantly higher than that for [t'], and for [p']. T-tests also reveal that there is a significant difference between [t'] and [p']. As for unaspirated stops, it is apparent that the mean VOTs for both [t] and [k] are significantly longer than that for [p]. It should also be noted that the mean VOT value for [t] is slightly higher than that for [k], which does not support the general agreement that velar stops are usually produced with longer VOTs. With respect to the mean VOTs without subject A, it shows almost the same trend with the one with subject A but differs in two places. First, the mean VOT values for all English stops, especially for [p', t', k'], are lower when subject A is excluded. One explanation for this may be that subject A is the only English-Mandarin bilingual among the subjects, and this is also reflected in the distribution (see results below). Secondly, VOT values for [k] is slightly higher than that for [t], which conforms more closely to the general principle, although the difference between [t] and [k] does not reach significance.

**Table 5. Mean VOTs for English stops by British English speakers with and without subject A.**

	/p'/	/t'/	/k'/	/p/	/t/	/k/
English (with subject A)	67	76	91	11	26	24
English (without subject A)	62	73	86	11	22	24

As for English VOT distribution, Figure 2 indicates that VOT ranges for [p', t', k'] are from 22 ms to 117 ms, 48 ms to 105 ms, and 65 ms to 145 ms, respectively; while the ranges

for [p, t, k] are 7-18 ms, 13-68 ms, and 13-40 ms, in that order. Some of the higher values, such as the highest value for [p’], the top two values for [k’], and the values from 52 ms to 68 ms for [t], were produced by subject A. Generally, VOT distribution in the present study fits the ranges reported in both Lisker and Abramson who examined American English, and Docherty [Docherty 1992], who concentrated on British English (their findings are provided in Table 1).

### 4.3 Comparing Mandarin and English VOT

VOT patterns in Mandarin and English are compared in terms of the mean VOT values and the distribution patterns. Figure 3 presents the mean VOTs for Mandarin and English stops. Chinese speakers generally produce longer VOTs than English speakers do, especially for voiceless aspirated stops. A series of t-tests were implemented to examine if the differences among /p’, t’, k’/ in both languages reach significance.

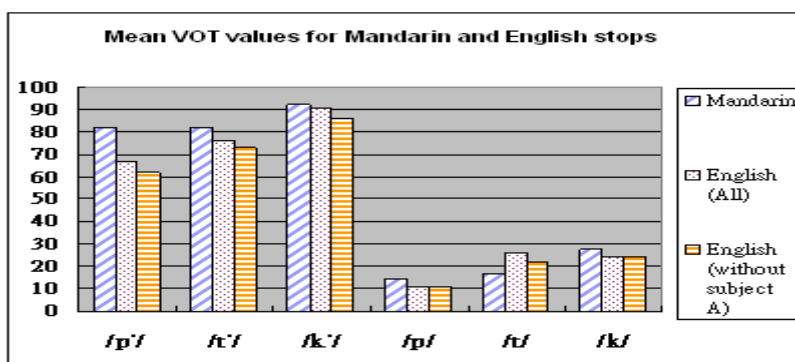
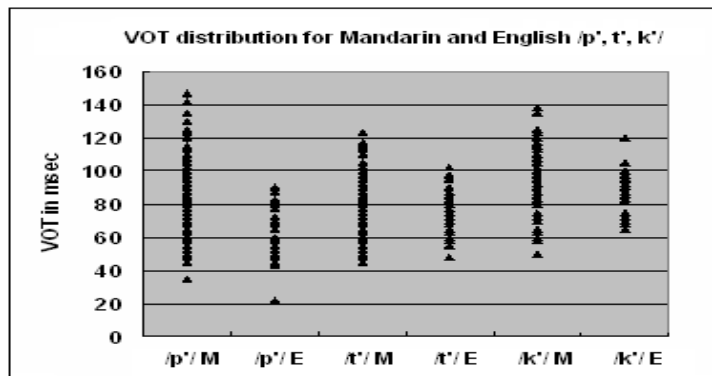


Figure 3. Mean VOTs for Mandarin and English stops by Chinese speakers and English speakers with and without subject A

Comparing the mean VOTs in Mandarin and English with subject A’s productions the results reveal that no comparisons reach significance although the figure indicates that VOT values for [p’, t’, k’] are longer in Mandarin than in English. However, when subject A is excluded from the group of English speakers, t-tests comparing Mandarin and English mean VOTs show significant results for all the aspirated stops although the differences between Mandarin and English VOT are subtle, not stark. It would thus be interesting to find out whether the L2 learners are aware of the subtle differences between the two languages and are capable of producing them authentically.

As Figure 3 shows, the mean VOT values for English [p’, t’, k’] are longer as the further back of the place of articulation. However, this does not apply to Mandarin [p’] and [t’]. The result accords with Rochet and Fei’s finding that the mean VOT durations for Mandarin [p’] and [t’] are close to each other, and [p’] always has a slightly higher value than [t’].

VOT distribution patterns for [p', t', k'] in Mandarin and in English are shown in Figure 4. Mandarin has higher VOT values than English; moreover, the ranges for Mandarin [p', t', k'] are wider than that for English ones. According to the definition for voiceless aspirated stops provided by Cho and Ladefoged [Cho and Ladefoged 1999], Mandarin occupies the 'highly aspirated' region along the VOT continuum while English falls into the 'aspirated' position. However, even if the two languages belong to the different categories, it does not mean that they occupy totally separate places along the VOT continuum. It can be clearly seen from the figure that the VOT ranges for Mandarin and English [p', t', k'] are considerably overlapped. The finding triggers a series of concerns about whether the native Chinese speakers will be aware of those subtle differences.



*Figure 4. VOT distribution patterns for Mandarin and English /p', t', k'/ produced by Chinese speakers (M) and English speakers (E) without subject A.*

## 5. Discussion

With respect to the comparison of VOT patterns in Mandarin and English, the results of this study corroborate the claim that significant differences exist between these two languages, especially for voiceless aspirated stops. It is known that all Mandarin and English stops are phonetically voiceless, and aspiration is becoming a common way of distinguishing homorganic pairs. According to Lisker and Abramson's categorization [Lisker and Abramson 1964], both Mandarin and English belong to the two-category group of languages, for which VOTs range from 0 to +25 ms for unaspirated [p, t, k] and from +60 to +100 ms for aspirated [p', t', k']. As expected, Mandarin and English follow short lag vs. long lag VOT patterns. However, the results of the study by Rochet and Fei [Rochet and Fei 1991] show that this classification is too general to observe the subtle differences between Mandarin and English. It is uninformative to compare unaspirated stops in the two languages as they occupy the same region along the VOT continuum and have close mean VOTs. As for aspirated stops, it may be

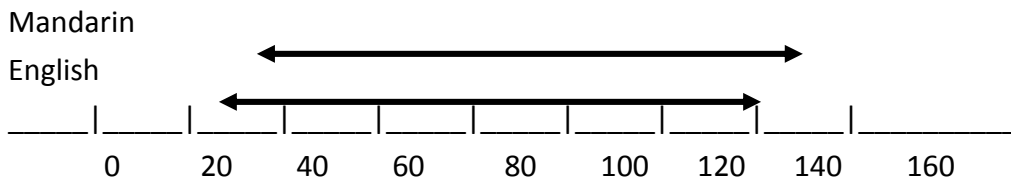
noted that Mandarin [p', t', k'] often have higher VOT values than English [p', t', k']. The findings showing higher mean VOTs in Mandarin are generally consistent with those reported by Rochet and Fei. However, the values obtained in this present study are slightly lower than Rochet and Fei because this experiment uses disyllabic words.

Aware that the three universal categories (*i.e.* long lead, short lag, and long lag) are not fine enough to distinguish the different regions along the long-lag VOT continuum, Cho and Ladefoged [Cho and Ladefoged 1999] suggest using four categories to define voiceless stops according to their degree of aspiration. In light of Cho and Ladefoged's categorization, it is clear that English [p', t', k'] fall into the 'aspirated' category with VOTs ranging from 50 ms to 90 ms. As for Mandarin, as shown in Table 6, all the VOT values of Mandarin [p', t', k'] in Rochet and Fei are over 90 ms, while in the present study the VOT values of the velar [k'] is over 90 ms. Therefore, Mandarin [p', t', k'] should be generally categorized as 'highly aspirated.' However, in this present study, Mandarin [p'] and [t'] might not fall into the 'highly' aspirated category. The results of the present study agree with Cho and Ladefoged's claim that more than three categories for aspirated stops need to be taken into consideration.

**Table 6. Mean VOT values (in ms) for Mandarin voiceless aspirated stops**

	Rochet and Fei 1991	The present study
	monosyllables	disyllables
p'	99.6	82.2
t'	98.7	81.9
k'	110.3	92.1

The study by Cho and Ladefoged used only VOT means for categorization. However, distribution is even more important. Figure 5 provides a simplified schematic representation of the places that Mandarin and English stops occupy along the VOT continuum. As shown in the figure, Mandarin [p', t', k'] occupies wider ranges and higher values than English. Moreover, it is difficult to draw a line between the two sets of VOT values for Mandarin and English [p', t', k'] due to the considerable overlap between the two distributions.



**Figure 5. Schematic representation of VOT ranges for English and Mandarin [p', t', k'] along the VOT continuum.**

Knowing that there is a great deal of overlap, it is interesting to find out whether the differences between Mandarin and English reach significance. The results of the present study show that all the differences for [p', t', k'] in both languages reach significance. It may thus imply that Mandarin and English occupy the same place along the VOT continuum for voiceless unaspirated stops, while the two languages belong to different categories for voiceless aspirated stops. Further studies are required using a greater number of monolingual subjects and examining [p', t', k'] in different contexts (in isolation and in sentences) in order to describe the VOT patterns of the Mandarin stop categories more accurately.

## 6. Conclusion

With the purpose of comparing voice onset time patterns in Mandarin and English as well as categorizing Mandarin stops by voice onset time, two general conclusions may be drawn from the present study. First, it is found that VOT patterns in the two languages are similar but not completely identical. Voiceless unaspirated [p, t, k], stop realizations in both languages occupy the same range: the short lag region along the VOT continuum. However, for voiceless aspirated [p', t', k'], Mandarin seems to fall into the 'highly aspirated' region along the VOT continuum, while English falls into the 'aspirated' region. The findings obtained in this study conform to the result reported by Rochet and Fei [Rochet and Fei 1991]. Moreover, the discrepancies between Mandarin and English are subtle due to the considerable overlap. Additional research with more monolingual subjects and a more natural setting (*i.e.* to obtain natural speech rather than elicited word lists) will be needed in order to accurately pinpoint the VOT patterns of Mandarin stops.

As for Mandarin stops categorization, the results suggest that the classification presented by Cho and Ladefoged [Cho and Ladefoged 1999] is more suitable than the three-way categorization [Lisker and Abramson 1964], especially for the language whose voicing contrast is the aspiration. Moreover, to discuss the distinction of voicing contrast, VOT is not the only parameter. Some researchers suggest other acoustic cues which share the equivalent importance with voice onset time to distinguish voicing contrast. These should be taken into account in future studies. Furthermore, two critical issues should be further examined: the irregular tendency for Mandarin [p', t', k'] and the underlying reasons for the differences between Mandarin and English VOT. The former can be discussed involving some possible factors, such as tones, vowel context, and place of articulation, while the latter can be detail explained by many aspects, for instance, different phonetic features or sound system in the two languages. Some researchers have claimed that tones play a role of being associated with VOT values [Liu *et al.* 2008]. The test stimuli used in the present experiment are not with the same tone; therefore, future studies should take it into consideration as well.

## References

- Bohn, O.S. and J.E. Flege, "Perceptual Switching in Spanish/English Bilinguals," *Journal of Phonetics*, 21(3), 1993, pp. 267-290.
- Caramazza, A., 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*, 54, 1973, pp. 421-428.
- Cho, T and P. Ladefoged, "Variation and Universals in VOT: Evidence from 18 Languages," *Journal of Phonetics*, 27, 1999, pp. 207-229.
- Docherty, G.J., *The Timing of Voicing in British English Obstruents*, Foris, New York, 1992.
- Fellbaum, M.L., "The Acquisition of Voiceless Stops in the Interlanguage of Second Language Learners of English and Spanish," *Proceedings of the 4<sup>th</sup> International Conference on Spoken Language Processing*, 3, 1996, Philadelphia, PA, pp. 1648-1651.
- Fischer-Jorgensen, E., "Acoustic Analysis of Stop Consonants," *Miscellanea Phonetica*, 1954, pp. 42-59.
- Flege, J.E., "Phonetic Approximation in Second Language Acquisition," *Language Learning*, 30(1), 1980, pp. 117-134.
- Flege, J.E. and W. Eefting, "Production and Perception of English Stops by Native Spanish Speakers," *Journal of Phonetics*, 15, 1987, pp. 67-83.
- Flege, J.E. and R.M. Hammond, "Mimicry of Non-Distinctive Phonetic Differences Between Language Varieties," *Studies in Second Language Acquisition*, 5(1), 1982, pp. 1-17.
- Keating, P.A., W. Linker, and M. Huffman, "Patterns in Allophone Distribution for Voiced and Voiceless Stops," *Journal of Phonetics*, 11, 1983, pp. 277-290.
- Keating, P.A., "Phonetic and Phonological Representation of Stop Consonant Voicing," *Language*, 60, 1984, pp. 286-319.
- Khattab, G., "VOT Production in English and Arabic Bilingual and Monolingual Children," *Leeds Working Papers in Linguistics and Phonetics*, 8, 2000, pp. 95-122.
- Klatt, D.H., "Voice Onset Time, Frication, and Aspiration in Word-Initial Consonant Clusters," *Journal of Speech and Hearing Research*, 18, 1975, pp. 686-706.
- Lisker, L. and A.S. Abramson, "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements," *Word*, 20, 1964, pp. 384-422.
- Lisker, L. and A.S. Abramson, "Some Effects of Context on Voice Onset Time in English Stops," *Language Speech*, 10, 1967, pp. 1-28.
- Liu, H., M. Ng, M. Wan, S. Wang and Y. Zhang, "The Effect of Tonal Changes on Voice Onset Time in Mandarin Esophageal Speech," *Journal of Voice*, 22(2), 2008, pp. 210-218.
- Port, R.F. and R. Rotunno, "Relation Between Voice-Onset Time and Vowel Duration," *Journal of the Acoustical Society of America*, 66, 1979, pp. 654-662.
- Riney, T.J. and N. Takagi, "Global Foreign Accent and Voice Onset Time Among Japanese EFL Speakers," *Language Learning*, 49(2), 1999, pp. 275-302.

- Rochet, B.L. and Y. Fei, "Effect of Consonant and Vowel Context on Mandarin Chinese VOT: Production and Perception," *Canadian Acoustics*, 19(4), 1991, p. 105.
- Rochet, B., T. Nearey and M. Munro, "Effects of Voicing Place and Vowel Context on VOT for French and English Stops," *Journal of the Acoustical Society of America*, 81, 1987, S65.
- Shimizu, K., A Cross-Language Study of Voicing Contrasts of Stop Consonants in Asian Languages, PhD thesis, University of Edinburgh, 1990.
- Weismer, G., "Sensitivity of Voice-Onset-Time (VOT) Measures to Certain Segmental Features in Speech Production," *Journal of Phonetics*, 7, 1979, pp. 197-204.



## Appendix A

Table A.1. Inventory of Mandarin stops followed by two high vowels /i/ and /u/, and a low vowel /a/.

	/p/	/t/	/k/	/p'/	/t'/	/k'/
/i/	逼迫 pi po	低聲 ti sheng	/	霹靂 p'i li	踢球 t'i qiu	/
/a/	八年 pa nian	答應 ta ying	嘎嘎 ka ga	趴下 p'a xia	踏過 t'a guo	卡鎖 k'a suo
/u/	布丁 pu ding	杜葳 tu wei	故宮 ku gong	撲倒 p'u dao	土匪 t'u fei	苦苓 k'u ling

Table A.2. Inventory of English stops followed by two high vowels /i/ and /u/, and a low vowel /a/.

	/p/	/t/	/k/	/p'/	/t'/	/k'/
/i/	beetle	decent	/	peeling	teacup	/
/a/	bath tub	darling	gagging	passion	tackle	castle
/u/	booting	duvet	google	poodle	tooth paste	cooling

