Crosslinguistic Influence on VOT Spectrum: A Comparative Study on English, Mandarin and Min

Jarry Chia-Wei Chuang
Department of English
National Chengchi University
Taipei, Taiwan
cwchuang.academia@gmail.com
ORCID: 0000-0002-3029-4463

Abstract

Crosslinguistic comparison of VOT has indicated linguistic transfer of voicing and aspiration contrasts in many languages. Mandarin has clear aspiration contrasts for voiceless stops, while Min presents another complicated VOT pattern, where voicing and aspiration contrasts are involved. The present study makes a crosslinguistic comparison between languages with voicing and aspiration contrasts as well as the potential linguistic transfer of VOT in English contexts from Mandarin and Min. There are three subject groups, including American English natives and Mandarin-Min bilinguals with different levels of Min-fluency. Mandarin-Min bilinguals have more aspirations and higher VOTs for aspirated voiceless stops than English natives. They also present two surfaces for English underlying voiced stops, voiced and unaspirated voiceless. Different levels of Min fluency are found to influence the tendencies towards voiced or unaspirated voiceless representations of English voiced stops. The overall finding presents a clear crosslinguistic influence on VOT patterns.

Keywords: Voice onset time (VOT), crosslinguistic, English, Mandarin, Min.

1 Introduction

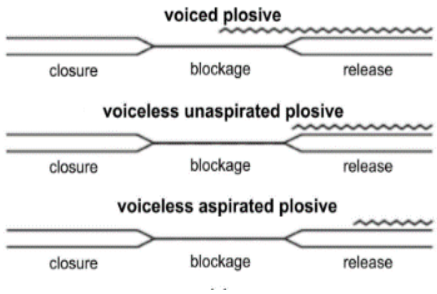
Researchers have been long explored phonetic and phonological acquisition in Mandarin-speaking ESL and EFL contexts (Chao & Chen, 2008; Chuang, 2021; Liu, 2017; among others). In Taiwan, Mandarin is the dominant language and English is the first foreign language, which has been taught from primary education to tertiary education. Taiwan EFL learners have been reported to have several phenomena of linguistic transfer from Mandarin to English, one of which is concerned with voicing contrast in stops. Previous studies on crosslinguistic influence have only taken Mandarin as the primary variable (Chao & Chen, 2008), while Crosslinguistic influences should be carefully noted. Taiwan is a multilingual community and most of the Taiwanese people are at least bilingual, though their L2 fluency may, subtly or divergently, differ from one to another. In Taiwan they are likely to acquire Min, Haka, or Austronesian languages; among all, Mandarin-Min bilinguals is the majority. To better capture the linguistic transfer of VOT, the paper will take different linguistic backgrounds and experiences (especially Mandarin-Min bilingualism) into account, revisiting crosslinguistic influences on VOTs in Taiwan English contexts via acoustic analysis.

2 Literature Review

2.1 VOT

Voice onset time (VOT) has led to phonetic studies in a vast number of languages. VOT marks the release of obstruction as well as reflects laryngeal vibration. Though its reliability for voicing distinction has been doubted in intervocalic or word-final positions (Docherty, 2011), it is still convincing that VOT can help identify voicing contrast in syllable-initial positions. VOTs show a categorical pattern for voicing and aspiration contrasts, in which VOT patterns can be classified into three possibilities (Lisker & Abramson, 1964), as shown in **Figure 1**.

Figure 1. Phonetic VOT patterns of plosives



The three-way distinction includes: (1) negative VOT (long lead), in which vocal folds vibrate far prior to the release of obstruction; (2) zero VOT (short lag), in which laryngeal vibration almost coincides with the unblocking of obstruction, often briefly delayed; (3) positive VOT (long lag), of which the occurrence is based on time priority of unobstructed airflow over the vibration. The three-way VOT patterns correspond to the voicing and aspiration of, particularly, stop consonants: (1) negative VOT for voiced stops, (2) zero VOT for voiceless unaspirated stops, and (3) positive VOT for voiceless aspirated stops.

The categorization is also be applied to the phonological analysis. [±voice] and [±spread glottis] feature in the linear phonological framework. Under Optimality Theory (OT), constraints in markedness and faithfulness adopt dichotomized judgments in voicing and aspiration as well.

2.2 VOT Spectrum

As the categorization seems well constructed for VOT, crosslinguistic findings reveal the deficiencies of the three-way VOT categorization and of the binary distinction in voicing and aspiration. In bilingual/multilingual contexts, speakers can produce divergent VOT values in the same category. To well account for the crosslinguistic evidence, Cho and Ladefoged (1999) proposed that VOT patterns should be better presented in a spectrum.

Aside from crosslinguistic influences on VOT, place of articulation and vowel contexts have been proven influential for VOT values in the same category. According to the aerodynamics resulting from jaw movements, the production of velar stops may contract the supraglottal cavity. It requires longer VOT, as it recovers from the formation of obstruction (Cho & Ladefoged, 1999). As for the surrounding vowels, tenseness and height of vowels might also contribute to the increase in VOT values (Klatt, 1975; Port & Rotunno, 1979; Weismer, 1979). So, a fine inspection of VOT values is thus required when we explore the voicing and aspiration contrasts, which has been accordingly considered in the experiment design.

2.3 Phonological & Phonetic Comparison

Normally, VOT patterns have corresponding phonological inventories of word-initial stops in the underlying representation (UR). Voicing and aspiration contrasts in English, Mandarin, and Min are differently distributed in UR. Crosslinguistic consonantal distributions of voicing and aspiration lead to a phonological comparison in **Table 1**.

	English	**Mandarin**	**Min**
Aspirated Voiceless		/p^h, t^h, k^h/	/p^h, t^h, k^h/
Unaspirated Voiceless	/p, t, k/	/p, t, k/	/p, t, k/
Voiced	/b, d, g/		/b, d, g/

Table 1. Phonological representation of word-initial contrasts in voicing and aspiration

Phonological representation may not be the final output, for aspiration and phonetic realization rule, as illustrated in **Figure 2.** Word-initial stops in English can phonologically be voiced /b, d, g/ or voiceless /p, t, k/ in UR. In SR, voiceless stops in word-initials become aspirated. Besides, underlying voiced stops in English have been reported to have two representations in SR, on the basis of the phonetic realization rule. One of them is voiced stops with vibration delays in VOTs, and the other is unaspirated voiceless stops with 15-20-ms vibration delays in VOTs. Among two surfaces of English voiced stops, the latter is the majority and greatly controls the mean VOT value. Phonetically speaking, English stops in word-initials mostly present phonological voicing contrast by aspiration in SR (i.e., phonetic difference between Zero VOTs and Positive VOTs).

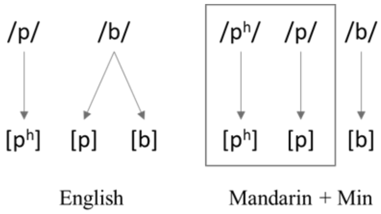


Figure 2. Phonetic realization of word-initial stops

In Mandarin and Min, aspiration contrasts of voiceless stops phonetically and phonologically play the major function of VOT. Mandarin has voiceless (un-)aspirated stops only, so there are no voicing contrasts. As for Min, the phonological inventory is sophisticated, with three categorized voicing and aspiration contrast phototactically acceptable in word-initial positions. Phonological patterns of onset plosives vary between English, Mandarin and Min, in which crosslinguistic performances are expected to display a wide array of patterns.

2.4 VOTs of Mandarin and English

In Chao and Chen's (2008) study, Mandarin-English VOT patterns have been comparatively examined. Both Mandarin and English VOTs fit the dichotomized classification (short lag vs. long lag), with aspiration contrasts in outputs only. Further visiting the voiceless aspirated stops, they indicate that VOTs in two languages are unidentical, in which VOTs of [p^h, t^h, k^h] in Mandarin are higher than those in English, which also provides evidence that VOT is better presented in a spectrum rather than a three-way distinction. Given the results, it is intriguing to see if crosslinguistic influence can trigger a linguistic transfer of VOTs from Mandarin to English.

3 Method

3.1 Subjects

Subjects were limited to be English natives or Mandarin-Min Bilinguals, with their ages ranging from 18 to 31 (*M* = 21.35; *SD* = 3.26).; in total, there have been 11 American English natives (5M; 6F) and 31 Mandarin-Min bilingual speakers (16M; 15F) in participation with the experiment. All the Mandarin-Min bilinguals were Taiwan Mandarin natives. Their Min fluency has been preliminarily self-identified and secondarily validified by 3 linguistically-trained Min natives in a 5-minute speaking test. Their Min fluency was scored with a Likert scale, from 1 (low) to 5(high) (M_L = 1.76; M_H = 4.67). Participants would be excluded as Min fluency was around 3 (intermediate) or presented a distinct mismatch between self-identified results and speaking checks by Min natives. There were 17 subjects with low fluency in Min (8M; 9F) and 14 subjects being fluent in both Mandarin and Min (7M; 7F). No speech disorders or diseases have been found.

Mandarin-Min transfer should be particularly noted here. Mandarin is much more dominant in Taiwan than Min. Here, we define Mandarin as their first language and Min as their second language with intimate language contact. Most of the linguistic transfers are Mandarin-to-Min, while the increase of Min fluency in Mandarin-Min bilinguals may lessen the transfer. High Min-fluency speakers may also have more linguistic habits brought from Min to other languages.

3.2 Stimuli

Word-initial stops were the major focus of the present study, as shown in **Table 1**. A total of 60 English words began with voiced and aspirated voiceless plosives and were adopted as the experimental materials for all the subjects. Mandarin-Min bilinguals were asked to read out extra 60 Mandarin words and 90 Min words, which were designed as disyllabic words for disambiguation of single-word senses. Tokens would be 10 words per stop in a language. Vowel contexts followed by target stops have been set with 5 tense high vowels and 5 lax low vowels for a stop. In total, 9 American English natives produced 540 English tokens. As for Mandarin-Min bilinguals, low Min fluency subjects produced 2550 tokens high Min fluency subjects produced 2100 tokens. It should be noted that 3 English tokens, 2 Mandarin tokens, and 7 Min tokens were reported to be invalid and have been excluded, with neglectable data loss.

3.3 Procedure

Before participating in the experiment, subjects had a self-evaluation of Min fluency and advanced evaluation by Min natives. All the subjects were then asked to read out English tokens in random order. Mandarin-Min bilinguals would further read Mandarin and Min tokens to check the VOT patterns in bilingual conditions.

3.4 Data Analysis

Acoustic data from the experiment has been imported to Praat 6.1.50 (Boersma, 2006) for analysis. VOTs were measured with waveforms and spectrograms, based on the interval between the release burst and the glottal vibration, as marked in **Figure 3**. The red is marked for the measurement of VOT.

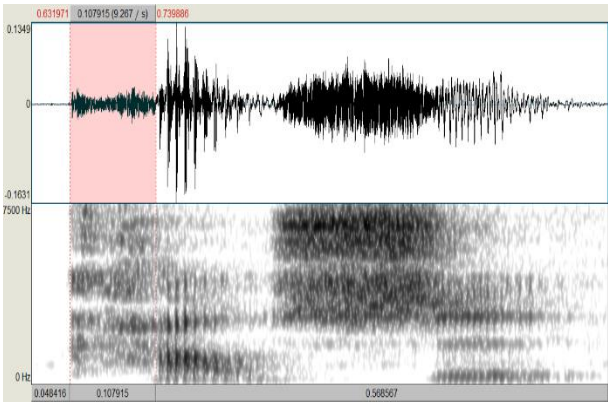


Figure 3. Acoustic analysis of *kǎo shì* ‘exam’ in the waveform and spectrograph.

4 Results & Discussion

4.1 Mean VOT

VOT patterns in Mandarin, Min and English contexts are presented as follows, along with the comparison between subjects with low (L) and high (H) Min fluency as well as English natives (E).

4.1.1 VOTs in Mandarin

For Mandarin VOT patterns, 3 aspirated voiceless stops /p^h, t^h, k^h/ and 3 unaspirated voiceless stops /p, t, k/ have been examined in **Table 2**.

	/p^h/	**/t^h/**	**/k^h/**	**/p/**	**/t/**	**/k/**
L	87.6	84.1	93.2	16.3	13.6	25.4
H	89.2	85.7	97.3	11.9	16.9	28.7
L+H	88.3	84.8	95.1	14.3	15.1	26.9

Table 2. Mean VOT values of Mandarin-Min bilinguals in Mandarin contexts

Mean VOT values show stops in Mandarin contexts contribute more to positive VOTs. We figure out no distinct divergence of VOT values between subjects with low and high Min frequency. In average, VOT values of aspirated voiceless stops are around 90 [$\bar{X}_1$(p^h, L+H) = 88.3; $\bar{X}_1$ (t^h, L+H) = 84.8; $\bar{X}_1$ (k^h, L+H) = 95.1]. As for unaspirated voiceless stops, the average VOT values are positive as well [$\bar{X}_1$ (p, L+H) = 14.3; $\bar{X}_1$ (t, L+H) = 15.1; $\bar{X}_1$ (k, L+H) = 26.9]. Mandarin VOT patterns generally show Mandarin aspirated voiceless stops have strong aspiration, with a long delay of glottal vibration.

4.1.2 VOTs in Min

Min phonology permits three major kinds of word-initial distributions in the VOT spectrum, including 3 aspirated voiceless stops /p^h, t^h, k^h/, 3 unaspirated voiceless stops /p, t, k/, and 3 voiced stops /b, d, g/. VOT patterns in Min contexts are shown in **Table 3**.

	/p^h/	**/t^h/**	**/k^h/**	**/p/**	**/t/**	**/k/**
L	99.6	97.1	101.7	11.7	10.6	23.2
H	102.3	103.8	111.9	8.2	9.3	27.1
L+H	100.8	100.1	106.3	10.1	10.0	25.0

	/b/	**/d/**	**/g/**
L	-97.1	-92.4	-133.4
H	-134.7	-119.5	-157.9
L+H	-114.1	-104.6	-144.5

Table 3. Mean VOT values of Mandarin-Min bilinguals in Min contexts

In Min, aspirated voiceless stops have positive VOTs, unaspirated voiceless stops have nearly Zero VOTs, and voiced stops have negative VOTs. Mean VOTs of aspirated voiceless stops falls on around

100 ms [$\bar{X}_2$(p^h, L+H) =100.8; $\bar{X}_2$ (t^h, L+H) =100.1; $\bar{X}_2$ (k^h, L+H) =106.3], and those of unaspirated voiceless are around 10-25 [$\bar{X}_2$(p, L+H)=10.1; $\bar{X}_2$ (t, L+H)=10.0; $\bar{X}_2$ (k, , L+H)=25.0]. VOTs of voiced stops in Min show the burst releases are much earlier than glottal vibration [$\bar{X}_2$(b, L+H)= -114.1; $\bar{X}_2$ (d, L+H) = -104.6; $\bar{X}_2$ (g, L+H)= -144.5]. The VOT patterns of Min are distinguishable for the clear three-way distinction.

4.1.3 VOTs in English

English phonology has only voicing contrasts, while it phonetically allows aspirated voiceless stops for underlying voiceless stops and unaspirated voiceless stops for voiced stops to appear word-initially. Crosslinguistic VOT patterns offer a well comparison between phonological representations and phonetic realization, as presented in **Table 4**.

	/p/	/t/	/k/	/b/	/d/	/g/
E	49.5	55.7	77.3	6.1	8.1	20.7
				-52.2	-70.5	-66.3
L	89.6	92.0	99.3	14.1	13.7	23.7
				-100.54	-95.9	-111.8
H	99.8	99.5	109.4	15.5	14.3	19.7
				-122.5	-125.1	-130.6
L+H	94.2	95.4	103.9	14.8	14.1	21.7
				-115.4	-115.1	-121.9

Note: Word-initial /p, t, k/ are [p^h, t^h, k^h]; VOTs of /b, d, g/ can be presented in short lag or long lead for phonetic realization.

Table 4. Mean VOT values in English contexts

Regarding aspirated voiceless stops for underlying voiceless stops, VOTs of English natives have vibration delays of around 50-80 ms [$\bar{X}_3$(p, E)=49.5; $\bar{X}_3$ (t, E)=55.7; $\bar{X}_3$ (k, E)=77.3] and Mandarin-Min bilinguals present much longer VOTs, up to 90-110 ms in average [$\bar{X}_3$(p, L+H) =94.2; $\bar{X}_3$ (t, L+H) =95.4; $\bar{X}_3$ (k, L+H) =103.9].

As for voiced stops [b, d, g], acoustic data has presented an intricate pattern. English natives present Zero VOTs [$\bar{X}_3$(b, E_1)= 6.1; $\bar{X}_3$ (d, E_1)= 8.1; $\bar{X}_3$ (g, E_1)=20.7] as well as negative VOTs (in the minority) [$\bar{X}_3$(b, E_2)= -52.2; $\bar{X}_3$ (d, E_2)= -70.5; $\bar{X}_3$ (g, E_2)= $-$ 66.3]. Besides, Mandarin-Min bilinguals also have two types of VOTs for English voiced stops: Their mean VOT values for English voiced stops are about 15 ~ 25 ms [$\bar{X}_3$(b, L+H_1)= 14.8; $\bar{X}_3$ (d, L+H_1)= -14.1; $\bar{X}_3$ (g, L+H_1)= 21.7], and around -115 ~ -120 ms [$\bar{X}$3(b, L+H_2)=-115.4; $\bar{X}$3 (d, L+H_2)=-115.1; $\bar{X}$3 (g, L+H_2)=-121.9]. Mandarin-Min bilinguals' VOT patterns of English voiced stops are more complicated than what figures show in the table, which will be further examined in 4.2.

4.2 VOT Distribution of English voiced stops

Though **Table 4** seems to show that Mandarin-Min bilinguals' potential tendencies towards negative VOTs as they produce English voiced stops, mean VOT values do not provide sufficient cues for such accounts. Their distributions are actually divergent, which needs careful analysis of the VOT distributions.

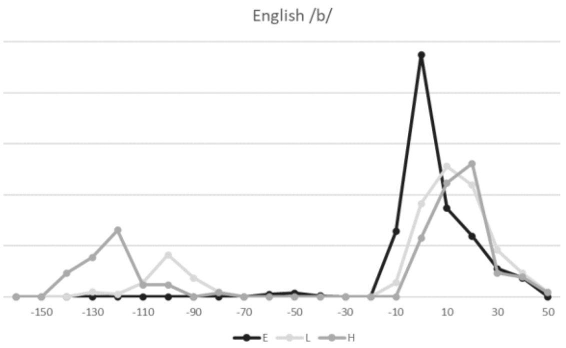


Figure 4. Surfaces of underlying /b/ in English

In **Figure 4**, English native speakers' VOT values for /b/ reach a peak at nearly 0 ms. They mostly produce Zero-VOT [p] for /b/. It should be noted that /b/ can [p] or [b] in the surface for phonetic realization rules, so a little number of subjects still present negative VOTs. In addition, Mandarin-Min bilinguals also produce /b/ in similar ways, but their occurrence rates show two divergent surface representations, which obviously differ from those produced by English natives. Mandarin-Min bilinguals' production of /b/ can show nearly Zero VOT (40 ~-10 ms) as well as negative VOTs centering on around -80 ~-150 ms, in which subjects with different Min fluency show the negative VOT peaks differently. High Min-Fluency subjects produce an earlier peak of negative VOTs (-120 ms) than low Min-fluency subjects (-100 ms), which reaches a statical significance ($p < 0.05$).

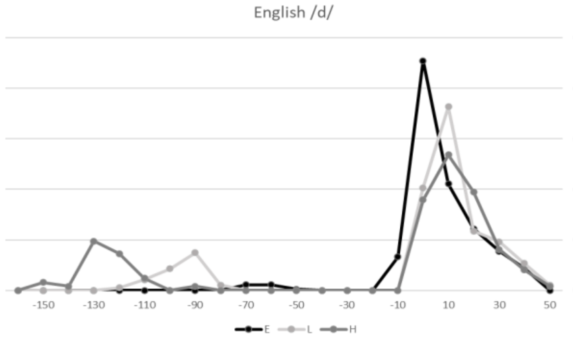


Figure 5. Surfaces of underlying /d/ in English

As for underlying /d/ in English, unaspirated voiceless [t] is the major surface. [d] can phonetically be the surface, so English natives still have little distribution of negative VOTs. Besides, in Mandarin-Min bilinguals' production of /d/, similar patterns with /b/ are found. They reach the negative VOT peaks at － 90 and － 130 ms, presenting a vast distribution of [b] around －80 ~ －140 ms. It is found that degree of Min fluency significantly influences the negative VOT peaks (p < 0.05).

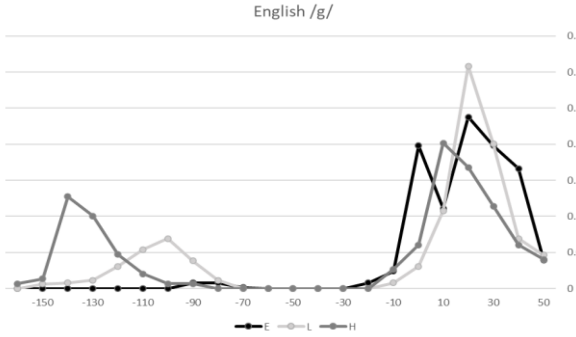


Figure 6. Surfaces of underlying /g/ in English

For the greater jaw movements, VOTs of [k] for underlying /g/ are higher than those of [p] for /b/ and [d] for /t/. Mandarin-Min bilinguals' VOT patterns of underlying /g/ thus present a more separate distribution between two surfaces, unaspirated voiceless stops [k] and voiced stops [g]. As to [g] for /g/, the peaks and the major distributions of negative VOTs by Mandarin-Min bilinguals are higher than [b] and [d]. The [d] distributions of high Min-fluency bilinguals reach the peak at －140ms; in comparison with low Min-fluency bilinguals, their intervals between the burst release and glottal vibration generally take shorter, around －90 ms. Data also reaches statistical significance (p < 0.05).

4.3 Crosslinguistic Comparison

In the study, crosslinguistic influences on VOTs are mainly shown in English contexts. Negative VOT patterns in English contexts show a complicated crosslinguistic influence across Mandarin, Min, and English. Different Min fluency levels further distinguish negative VOT patterns. Group L, in which Min is less dominant, shows a shorter negative VOT than Group H. For their L1, Mandarin, has no voiced stops in the inventory, their negative VOTs are not as longer/many as those produced by Group H. In general, low Min-fluency subjects have more linguistic transfers from Mandarin to English contexts.

Moreover, aspirated stops in English contexts also provide informative divergences in VOT values. Mandarin-Min bilinguals with high and low Min fluency present longer VOTs for [p^h, t^h, k^h], since Mandarin and Min contexts, in preference to English, are sensitive to aspiration contrasts. This finding corresponds with Chao and Chen's (2008) observation. Overall, English contexts offer comparative information about crosslinguistic influences on VOTs.

5 Conclusion

The present study conducts a crosslinguistic comparison between languages with voicing and/or aspiration contrasts. Results reveal the linguistic transfer of VOT appears mostly in English contexts. More aspiration, with higher VOT values, has been made by Mandarin-Min bilinguals for aspirated voiceless stops than by English natives, since stop contrasts are well constructed by aspiration in Mandarin and Min. Besides, Mandarin-Min bilinguals present two variations of English voiced stops, phonetically voiced and unaspirated voiceless. Different levels of Min fluency are found to influence speakers' tendencies. Low Min-fluency subjects produce more [p, t. k] for /b, d, g/ and shorter negative VOTs, as their dominant language, Mandarin, originally has no negative VOTs in the phonological inventory and phonetic realization. The findings generally demonstrate a clear crosslinguistic influence on VOT patterns.

Acknowledgments

The study has been supported by a grant from College of Foreign Languages & Literature, National Chengchi University. Besides, the author would like to thank Hsiu-Jung Li and Chih-I Chuang for scoring subjects' Min fluency.

References

Boersma, P. (2006). Praat: doing phonetics by computer. *http://www. praat. org/*.

Chao, K.-Y., & Chen, L.-m. (2008). *A cross-linguistic study of voice onset time in stop consonant productions.* Paper presented at the International Journal of Computational Linguistics & Chinese Language Processing, Volume 13, Number 2, June 2008.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of phonetics, 27*(2), 207-229.

Chuang, C.-W. (2021). *Mandarin Speakers' Acquisitions and Representations of Flapping in American English in An ESL Context: A Perception and Production Study*. Paper presented at the 2021 24th Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA).

Docherty, G. J. (2011). The timing of voicing in British English obstruents. In *The Timing of Voicing in British English Obstruents*: De Gruyter Mouton.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of speech and hearing research, 18*(4), 686-706.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*(3), 384-422.

Liu, D. (2017). The Acquisition of English Word Stress by Mandarin EFL Learners. *English Language Teaching, 10*(12), 196-201.

Port, R. F., & Rotunno, R. (1979). Relation between voice‐onset time and vowel duration. *The Journal of the Acoustical Society of America, 66*(3), 654-662.

Weismer, G. (1979). Sensitivity of voice-onset time (VOT) measures to certain segmental features in speech production. *Journal of phonetics, 7*(2), 197-204.