# Effects of Prosodic Focus on Voice Onset Time (VOT) in Chongming Chinese

**Yitian Hong**
The Chinese University of Hong Kong,
Shatin, NT, Hong Kong SAR, China
hongyt@link.cuhk.edu.hk

**Si Chen**
The Hong Kong Polytechnic University,
Kowloon, Hong Kong SAR, China
sarahchen@polyu.edu.hk

**Yike Yang**
The Hong Kong Polytechnic University,
Kowloon, Hong Kong SAR, China
yi-ke.yang@connect.polyu.hk

**Bei Li**
The Hong Kong Polytechnic University,
Kowloon, Hong Kong SAR, China
benita.li@polyu.edu.hk

## Abstract

Prosodic focus is phonetically realized by increasing intensity, extending duration, and expanding pitch rage of focused components (Xu et al., 2012). Previous studies have also found the effect of prosodic focus on enlarging two-way or three-way stop contrast by lengthening the VOT (voice onset time) of voiceless or aspirated consonants (e.g. Choi, 2003; Chen, 2011). The present study investigates the influence of prosodic focus on the realization of VOT of an under-studied language, Chongming Chinese. Twelve monosyllabic words were selected and embedded in carrier sentences with different discourse conditions: one baseline neutral focus condition and three focus conditions. Precursor questions were prepared to elicit production from native speakers of Chongming Chinese. Results showed the significant main effects of stop types and discourse conditions on VOT realization. VOTs were shortened in unaspirated and voiced stops in the on-focus condition, suggesting a different way of expanding the three-way difference in stops. VOT was also affected by other focus conditions, providing implications for the study of focus domain. The study suggests that VOT can serve as acoustic cue for stop contrasts in Chongming Chinese in different prosodic environment and contributes new data to the typology of prosodic focus study as well as stop contrast research.

Index Terms: prosodic focus, VOT, phonological contrast, Chongming Chinese

## 1 Introduction

Prosodic focus refers to the use of speech prosody to emphasize a specific part of an utterance (Xu et al., 2012). Research across world languages shows that prosodic focus is realized with various acoustic cues including raising intensity, elongating duration, and expanding pitch range of the focused elements (e.g., Jong, 2004; Chen, 2011; Alzaidi et al., 2019). Accumulative evidence has found post-focus realization, mostly refers to compressed pitch range after the focused elements (e.g., Chen et al., 2009; Lee and Xu, 2018; Xu, 2011). The reduction of intensity was also found in post-focus condition (Chen et al., 2009). Several studies also investigated the pre-focus region of the utterance. Pre-focus pitch compression was reported in a Japanese study in both information focus and contrastive focus without plausible explanation (Hwang, 2012). Tone 3 in Mandarin Chinese also demonstrated the increase of duration, intensity and raising pitch of pre-focused components (Lee, 2015). However, the observations of both post-focus and pre-focus encodings vary from languages to languages (e.g., Lee and Xu, 2018; Xu et al., 2012)

Voice Onset Time (VOT) has been defined as “the interval between the release of the stop and the onset of glottal vibration” (Lisker and Abramson, 1964: 389). It was found to be a relatively reliable acoustic measurement for differentiating phonemic categories (i.e., voicing, aspiration, and force of articulation) of stops in a language (Lisker and Abramson, 1964). Factors that are likely to affect



Pacific Asia Conference on Language, Information and Computation (PACLIC 33), pages 414-422, Hakodate, Japan, September 13-15, 2019

VOT include the following vowel, features of the stops, and the prosodic environment. VOT is sensitive to the duration of the following vowel (Ling and Liang, 2016). Tense vowels were more likely to be preceded by longer VOT and lax vowels by shorter VOT (Port and Rotunno, 1979). The place of articulation of stops also plays a role in VOT. The longest VOT is found in velar stops, while the difference of VOT between alveolar and bilabial stops varies across languages (Ling and Liang, 2016; Lisker and Abramson, 1964).

In addition to the place of articulation and the following vowel duration, previous studies also examined the effects of prosodic focus on VOT. In the study of English, Choi (2003) reported that both voiced and voiceless stops demonstrated an apparent increase of VOT under focus, particularly voiceless stops. Their difference was also enlarged, suggesting the enhancement of stop voicing contrast. However, Cho and McQueen (2005) reported different findings in Dutch. The VOT of voiced stops was increased under focus while the VOT of voiceless stops was shortened. Dutch has a shorter VOT in voiceless stops than English. The increase of VOT in English in an on-focus condition signaled the enhancement of a [+aspirated] feature, while the reduced VOT in Dutch indicated the strengthening of the [-aspirated] feature (Ling and Liang, 2016).

Similar studies were conducted in a few languages with three-way contrastive stops, especially in Shanghai Chinese, a Wu dialect. Chen (2011) found significant lengthening of VOT in aspirated stops under the influence of focus, but not in unaspirated stops and voiced stops. Ling and Liang (2016) reported a significant increase on VOT in both aspirated and voiced stops of Shanghai Chinese in an on-focus condition, which was explained by maximizing phonological contrast among the three stop types. Cross-linguistic differences on the encoding of voicing and aspiration contrast in focus conditions are observed. However, little is known about the realization of VOT in various focus conditions across languages. Therefore, the present study aims to contribute novel data to the study of prosodic focus and phonological contrast by investigating the encodings of VOT in a Chinese dialect under various focus conditions.

Chongming Chinese is a tonal language and also a variety of Wu dialect. It is spoken by people living in eastern China, including Chongming County, Haimen City, Qidong City, Shazhou County, etc. (S. Chen, 2014). Similar to other Wu dialects, the three-way contrasts in the onset stops are reported (i.e., voiceless aspirated, voiceless unaspirated, and voiced). Previous studies have shown the effects of focus on F0, duration and intensity in Chongming Chinese (Yang et al., 2018, 2019), indicating that Chongming Chinese is a language with noticeable encodings of prosodic focus. The current study will examine a new perspective of focus realization in Chongming Chinese to enrich the study of the Wu dialect.

Based on the above review, three research questions are raised:

1) What are the differences among the VOT of stop types in neutral focus condition and focus conditions?
2) What effects of focus on the VOT of target stops can be found?
3) Is manipulation of phonological contrast observed in the VOT of Chongming Chinese under focus conditions? If yes, how is it achieved?

Provided previous findings of the realization of prosodic focus in Chinese languages as reviewed above, our prediction is that there will be significant differences among three stop types between neutral focus and focus conditions in order to maximize their phonological contrast.

## 2 Methods

### 2.1 Subjects

Twelve Chongming Chinese native speakers (six males, six females), aged 38 to 57 (mean ± SD: 52.00 ± 4.53), were recruited for the current study. According to their self-reports, Chongming Chinese is their mother tongue and dominant language for daily communication. They have never received any formal musical training and none of them reported speaking, hearing or language difficulties.

### 2.2 Stimuli

Twelve monosyllabic words varying in tones and initial stop types were selected as stimuli, as is illustrated in Table 1. Only one vowel was embedded in the stimuli in order to control the effect of following vowel duration. The vowel /æ/ was selected because it was found to appear in most tones for every consonant onset (S. Chen, 2014).

| | t | $t^h$ | d |
|---|---|---|---|
| **T1 (55)** | tæ | $t^h$æ | |
| **T2 (24)** | | | dæ |
| **T3 (424)** | tæ | $t^h$æ | |
| **T4 (242)** | | | dæ |
| **T5 (33)** | tæ | $t^h$æ | |
| **T6 (313)** | | | dæ |
| **T7 (5)** | tæ | $t^h$æ | |
| **T8 (2)** | | | dæ |

**Table 1.** *Target Stimuli*

Only alveolar stop was adopted to control the effect of place of articulation.

These stimuli were embedded in carrier sentences where contexts and discourse conditions were manipulated [illustrated in (1)]. The tones of the syllables before and after the target stimuli were controlled by selecting two syllables respectively before (i.e., Part C) and after (i.e., Part D) the target stimuli. Four combinations were generated (i.e., /ɦuo313/ and /ɦnø24/, /ɕiɑ424/ and /ɦnø24/, /ɕiɑ424/ and / ɦnø24/, /ɕiɑ424/ and /ɦnø24/).

(1)

| | | |
|---|---|---|
| *ɦmei24ɦain24* | *ɦuo313* | *ɦuo313/ɕia424* |
| matchmaker | say | say/write |
| (A) | (B) | (C) |
| *TARGET* | *ɦnø24/tu55* | *ɦlɛ24tɕiɔ55kuœ55* |
| TARGET | difficultly/much | very |
| | (D) | (E) |

'The matchmaker said that she said/wrote TARGET far difficultly/more.'

Four discourse conditions were employed in the carrier sentences: neutral focus condition, pre-focus condition, on-focus condition and post-focus condition, which signaled the position of the target stimuli, as shown in Table 2.

| Discourse conditions | Carrier sentences |
|---|---|
| neutral focus | (A) (B) (C) TARGET (D) (E) |
| pre-focus | (A) (B) ***(C)*** TARGET (D) (E) |
| on-focus | (A) (B) (C) ***TARGET*** (D) (E) |
| post-focus | (A) (B) (C) TARGET ***(D)*** (E) |

**Table 2.** *Carrier sentences designed in different discourse conditions (The foci are in bold and italic)*

## 2.3 Procedure

All the subjects were recorded in a quiet room in Qidong City. One PC demonstrated stimuli in E-prime (Schneider et al., 2012) for subjects' reference and another PC were used for recording by Praat (Boersma and Weenink, 2001).

In the neutral focus condition, subjects were instructed to read the carrier sentences in natural and normal speech. In three focus conditions, precursor questions were prepared for eliciting responses (the carrier sentences in Table 2). There were 2304 total tokens produced (12 target stimuli * 4 contexts * 4 focus conditions * 12 speakers).

## 2.4 Data Analysis

The consonants of the target stimuli were manually segmented by one trained phonetician in Praat (Boersma and Weenink, 2001) and checked by the other phonetician. The VOTs were labeled from the point of the stop release to the onset of the second formant of the preceded vowels. The Praat script ProsodyPro (Xu, 2013) was used for extracting VOT values. 112 tokens were suspected as incorrect production (i.e., produced in incorrect tones or had abnormal VOT) and confirmed by a native speaker. There were excluded from further analysis.

By plotting all the VOT values, we observed apparent inter-speaker variations. To investigate factors that significantly affect the VOTs, we fitted a basic linear mixed effect model to VOTs with subject as random effect by adopting the 'lmerTest Package' (Kuznetsova et al., 2017) in R (R Core Team, 2013). By adding stop types, discourse conditions and their interaction as fixed effects one after another, we improved the model. Next, we examined the contribution of stop types and discourse conditions, respectively, by fitting the linear mixed effect model again. The above model is sufficient to interpret the results.

Furthermore, we calculated the differences between different stops types in VOTs across four discourse conditions and compared them by plotting. All the figures were plotted by the 'ggplot2' package (Wickham, 2016) in R.

| Fixed effect | Estimate | SE | *t* | *P* |
|---|---|---|---|---|
| **(Intercept)** | 18.222 | 1.700 | 10.718 | <0.001*** |
| **Stop type: unaspirated** | -6.376 | 1.141 | -5.586 | <0.001*** |
| **Stop type: aspirated** | 20.169 | 1.159 | 43.289 | <0.001*** |
| **Discourse condition: on-focus** | -3.078 | 1.160 | -2.653 | <0.01** |
| **Discourse condition: post-focus** | -3.649 | 1.162 | -3.140 | <0.01** |
| **Discourse condition: pre-focus** | -3.515 | 1.155 | -3.042 | <0.01** |

Signif. codes: ***p<0.001, **p<0.01, *p<0.05 (the same across the whole paper)

**Table 3.** *Linear mixed model of VOT (2192 observations) (Voiced stops and neutral focus as baseline)*

| | | **voiced vs. unaspirated** | **voiced vs. aspirated** | **aspirated vs. unaspirated** |
|---|---|---|---|---|
| Neutral focus condition | | $t$ = -5.461, $p$<0.001 | $t$ = 42.145, $p$<0.001 | $t$ = -47.86, $p$<0.001 |
| Focus conditions | pre-focus | $t$ = -3.276, $p$<0.01 | $t$ = 49.018, $p$<0.001 | $t$ = -52.90, $p$<0.001 |
| | on-focus | $t$ = -3.750, $p$<0.001 | $t$ = 45.341, $p$<0.001 | $t$ = -49.80, $p$<0.001 |
| | post-focus | $t$ = -2.971, $p$<0.01 | $t$ = 45.328, $p$<0.001 | $t$ = -48.75, $p$<0.001 |

**Table 4**. *t statistics and p-values in the linear mixed model of VOT in different discourse conditions*

## 3 Results

The VOTs extracted from the target stops were analyzed by a linear mixed effect model with subject as a random effect. The basic model was improved by adding Stop type ($\chi^2$=4066.7, Df=2, $p$<0.001) and Discourse condition ($\chi^2$=10.067, Df=3, $p$<0.05) as fixed effects. However, their interaction did not play a significant role ($\chi^2$=8.9065, Df=6, $p$=0.1789), indicating that contrast of three stop types remained identical in different discourse conditions. The significant results are reported in Table 3. From Table 3, it can be inferred that the VOTs of unaspirated stops are significantly different from voiced stops. The difference between aspirated stops and voiced stops also reached significance. In terms of the discourse condition, the difference of neutral focus vs. on-focus, neutral focus vs. post-focus, and neutral focus vs. pre-focus reached significance.

We first analyzed the difference of stop types in four discourse conditions respectively. Results showed that the effect of stop type is significant across discourse conditions (Table 4). Pairwise significant differences were found among voiced, unaspirated, and aspirated stops when they were in neutral focus sentences or as pre-focused, on-focused, and post-focused elements of focused sentences.

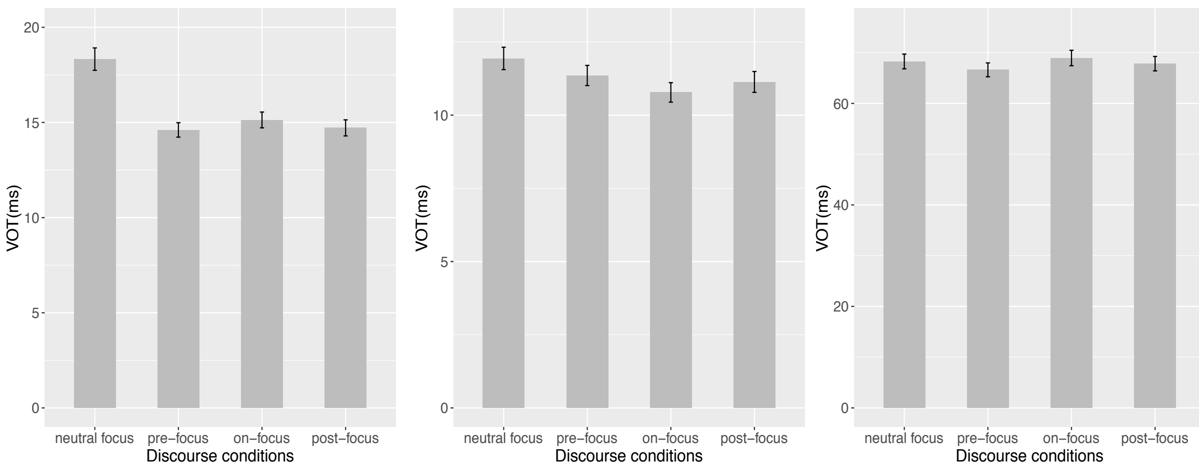
The results indicate that the differences among the stop types remains relatively stable in neutral focus condition and focus conditions. Three stop types were distinguishable from each other in VOT regardless of discourse conditions.

We then analyzed each stop type individually. The main effect of the discourse condition was found in the VOT of voiced stops, as analyzed below. The VOT of unaspirated stop also showed a significant difference between the neutral focus condition and the on-focus condition. No significant difference was found in the VOT of aspirated stop between neutral focus condition and any focus conditions.

When the voiced target syllable is in a pre-focus condition, its VOT is significantly shorter than in a neutral focus condition. The same significant differences were found in the on-focus vs. neutral focus and the post-focus vs. neutral focus condition (Table 5). The mean VOT of voiced stops in the on-focus condition decreases 17.54% compared to that in the neutral focus condition, while the mean VOT

| **Fixed effect** | **Estimate** | **SE** | ***t*** | ***p*** |
|---|---|---|---|---|
| **Intercept** | 18.2521 | 0.8791 | 20.763 | <0.001 *** |
| **pre-focus** | -3.6721 | 0.5878 | -6.247 | <0.001 *** |
| **on-focus** | -3.2018 | 0.5902 | -5.425 | <0.001 *** |
| **post-focus** | -3.6383 | 0.5911 | -6.155 | <0.001 *** |

**Table 5.** *Linear mixed model comparing VOT of voiced stop across discourse conditions (neutral focus condition as baseline)*




(a). *Voiced stops* (b). *Unaspirated stops* (c). *Aspirated stops*

**Figure1.** *Mean VOT of different stop types*

in pre-focus and post-focus conditions drops 20.12% and 19.93%, respectively, as shown in Figure 1(a). The figure indicates that when there is a focus before or after the target syllable, the VOT of the target voiced stop tends to be reduced more to signal the coming focus. When the focus is exactly the target syllable, the VOT also reduces, but to a lesser degree.

For unaspirated target stops, a significant difference was found between the VOT in the neutral focus condition and the on-focus condition ($t$= -2.450, $p$ <0.05*), suggesting that speakers reduced the VOT of the target syllable when it was the focus of the sentence. A marginally significant difference was found between the VOT in the post-focus condition and the neutral focus condition ($t$= -1.754, $p$= 0.0798). Figure 1(b) shows the mean VOT of unaspirated stops. Similar to voiced stops, the mean VOT of unaspirated target stops is shorter in all the focus conditions, among which, the VOT in the on-focus condition drops the most. Speakers tended to shorten the VOT of the target syllables to indicate that a focus was addressed in the sentence.

Aspirated target stop demonstrates a different pattern [Figure 1(c)]. The on-focus condition has the longest mean VOT (68.96ms), longer than the neutral focus condition. The VOT of aspirated stops in the post-focus condition is slightly shorter than that in the neutral focus condition, while the VOT in the pre-focus condition remains the shortest (66.62ms). Although the differences do not reach significance, when the target syllable is the focus, a trend for subjects to lengthen their VOT is observed. Subjects tended to emphasize the on-focus target aspirated stop by extending the VOT. It is also possible that speakers tried to differentiate an aspirated stop from the other two stop types by enlarging their VOT difference in the on-focus condition.

For further comparison, we calculated the mean differences among the three stop types in neutral focus condition and focus conditions by a two-two subtraction. The results are shown in Figure 2. The differences between aspirated and voiced stops increase in all the focus conditions compared to the neutral focus condition. On the contrary, the difference between unaspirated and voiced stops decreases in all the focus conditions. The difference between aspirated and unaspirated stops enlarges in the on-focus condition, while it reduces in the pre-focus condition, in comparison with the neutral

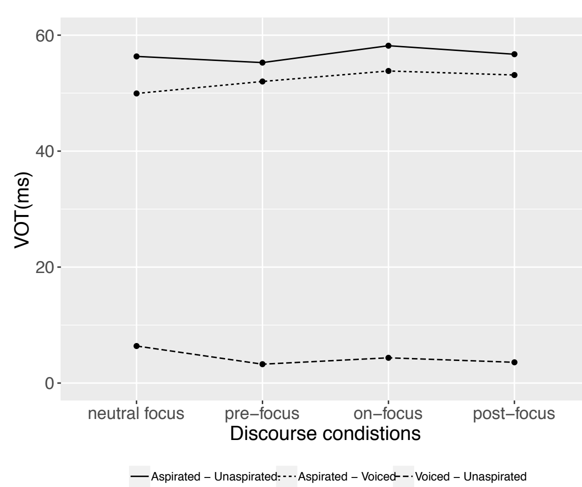



**Figure 2.** *Mean differences between stop types*

focus condition. The results suggest that, when a focus is indicated in the sentence, subjects are more likely to distinguish an aspirated target stop from unaspirated and voiced target stops. They reduced the distance of VOT between voiced and unaspirated target stops and increased the differences between aspirated and unaspirated and aspirated and voiced target stops. This tendency is more obvious when the target stops are the on-focused elements of the sentences.

The main findings of the current study are summarized as follows: 1) stop types show a significant effect on the VOT of target stops in neutral focus condition and three focus conditions; 2) in regards to discourse conditions, significant differences were found in the neutral focus vs. on-focus, the neutral focus vs. post-focus, the neutral focus vs. pre-focus condition of voiced stops and the neutral focus vs. on-focus condition of unaspirated stops; 3) compared to the neutral focus condition, the differences between VOT of aspirated and unaspirated and aspirated and voiced stops are enlarged in focus conditions, especially in the on-focus condition.

## 4 Discussion

The present study investigates the effects of stop types and prosodic focus on the VOT of Chongming Chinese. The results showed pairwise significant differences among VOTs of voiced, aspirated, and unaspirated stops in the neutral focus condition and the focus conditions. The VOT of voiced stops was significantly different between the neutral focus condition and the pre-focus/on-focus/post-focus conditions. For unaspirated stops, the difference between the VOT in the neutral focus condition and the on-focus condition also reached significance.

For Chongming Chinese, the VOT can distinguish the three-way contrast of stops. The difference of VOTs among stop types remained stable regardless of focus condition, which suggests that the manipulation of prosodic structure does not change the basic three-way distinction among stops in Chongming Chinese. The [+voiced] and the [+aspiration] features of stops are distinguishable from the measurement of VOT.

The VOT of voiced stops lies between the VOTs of aspirated stops and unaspirated stops in the current study, which may reveal its breathy nature, as indicated in Z. Chen (2014). It is commonly seen in Wu dialects that the voiced stops are pronounced in the manner of a weak voiceless onset proceeding a phonated breathy vowel (Ibid.). The VOTs of voiced stops were significantly longer than unaspirated stops in all the discourse conditions, indicating that VOT has the possibility of acting as a reliable acoustic cue for differentiating voiced stops and unaspirated stops in Chongming Chinese. It contradicts the general claim that in the Wu dialect, the voice feature and F0 shown in the following vowels are the major acoustic cues for distinguishing stops due to the similarity of VOTs between unaspirated and voiced stops (Ling and Liang, 2016; Z. Chen, 2014).

When the target syllables were on-focused, the VOTs of voiced and unaspirated stops were significantly shorter than VOTs in neutral focus condition, while the VOT of aspirated consonants is lengthened in the on-focus condition but did not reach significance. By doing so, the differences between unaspirated and aspirated stops and voiced and aspirated stops were enlarged. Our findings are consistent with the findings in other languages in that in the positions with prosodic emphasis, the phonetic components tend to be realized with the aim of enlarging phonological contrast (Chen, 2011; Cho and Keating, 2001; Choi, 2003). VOT is used for measuring both voicing and aspiration (Lisker and Abramson, 1964). For aspirated and unaspirated stops, the [-aspiration] feature is enhanced via the shortening of the VOT in the unaspirated stop to emphasize its difference from the aspirated stop. For voiced and aspirated stops, the [+voice] feature is strengthened in the voiced stop to reinforce its contrast with the aspirated stop. Therefore, the lexical contrast can be enhanced under focus. Moreover, the goal of spreading new information by providing focus in the utterance can be achieved.

Part of the results is not in line with the findings of Shanghai Chinese (Chen, 2011; Ling and Liang, 2016), in which aspirated stops and voiced stops increased the VOT significantly to signal the on-focus condition, while unaspirated stops remained stable. Because the prosodically conditioned lengthening effect (Cho and McQueen, 2005) is not apparent in the VOT of aspirated stops, Chongming Chinese speakers have to shorten the VOTs of unaspirated and voiced stops to compensate for maximizing the three-way stop contrast.

To find a possible explanation for the above inconsistent findings, we conducted a search and calculation of the *Chongming Fangyan Cidian* ('Dictionary of Chongming Dialect') (Zhang, 1993). We found that there is a total of 268 syllables with unaspirated stops, 174 syllables with voiced stops and 118 syllables with aspirated stops. All these syllables have independent lexical meanings in Chongming Chinese. The number of syllables with aspirated stops is smaller than the number of syllables with the other two stop types. It indicates that the frequency of appearance of voiced and unaspirated stops is much higher than aspirated stops in Chongming Chinese. Due to the lower frequency of occurrence of aspirated stops, speakers have less chance to practice manipulating their VOTs in different prosodic environments. They tend to rely more on adjusting the VOTs of the other two stop types to maintain the contrast, which may also explain why the increase of VOTs in on-focused aspirated stops was not significant.

The performance of voiced stops in focus conditions is unexpected. All the focus conditions demonstrated significant difference from neutral focus condition, suggesting that the voiced stops may play an extremely important role in maintaining stop contrast in Chongming Chinese.

As we discussed above, voiced stops in Chongming Chinese are presumably breathy stops. Breathy phonation refers to a situation when the vocal folds are both opening and vibrating (Davenport and Hannahs, 2013). Due to the escape of air, the energy for vibration is reduced and thus the [+voice] feature is not robust (Ibid.). When a focus is addressed, it is likely that Chongming Chinese speakers tried to compress the breathy nature of voiced stops by controlling the escape of air from the vocal folds and the VOTs were thus affected. By compressing the aspiration of breathy voiced stops, the [+voice] feature is strengthened and thus the voiced stops can distinguish themselves in the on-focus condition.

Another interesting issue revealed in our study is that the contrast between voiced and unaspirated stops was reduced in the on-focus condition, which seems to contradict the phonetic contract enhancement finding. It is likely that in Chongming Chinese, it is less important to draw the difference between voiced and unaspirated stops. We calculated all the minimal pairs for stops in the *Chongming Fangyan Cidian* ('Dictionary of Chongming Dialect') (Zhang, 1993). We found 317 minimal pairs for syllables with aspirated stops and unaspirated stops. Except for the difference in initial consonants (/p/ vs. /$p^h$/, /t/ vs. /$t^h$/, /k/ vs. /$k^h$/), their vowels and tones remain the same. In a similar manner, we found only 51 minimal pairs for syllables with voiced stops and unaspirated stops (/p/ vs. /b/, /t/ vs. /d/, /k/ vs. /g/). Fewer minimal pairs suggest the possibility of sacrificing the contrast between voiced and unaspirated stops and adding more efforts in distinguishing aspirated stops.

This study also tried to examine the target stops in pre-focus and post-focus positions. The post-focused influence was witnessed in voiced stops with a significant drop in the VOT from the neutral focus condition. Unaspirated stops also showed a marginally significant reduction of VOT in post-focus positions compared to the neutral focus condition. Previous studies have found that in a Verb Phrase (VP), when the initial verb was focused, the other arguments inside the VP (i.e., the oblique and thematic arguments) received prominence as well (Jun et al., 2006; Jun, 2011). The on-focused verb had the most robust emphasis (Ibid.). Referring back to the carrier sentence (1), the target syllable is the object of a VP, *ɦuo313/ɕia424 TARGET* ('say/write TARGET'). When the verb *ɦuo313/ɕia424* is on-focused, it is possible that the focus domain extends to the whole VP. As a post-focused component, as well as an argument of the VP, the target syllable may receive a certain degree of emphasis. Therefore, it is not surprising that in the post-focus condition, voiced and unaspirated stops still showed significant reduction to enlarge the three-way contrast in stops. This finding is in line with previous study of duration and intensity range change in post-focus condition in Chongming Chinese (Yang et al., 2019).

The VOTs were shortened in pre-focused words across all types of onset stops, among which, the reduction in voiced stops reached significance. The distinction between the voiced and unaspirated stops and the aspirated and unaspirated stops are reduced. Thus, it is hypothesized that speakers reduced the three-way stop contrast to differentiate the pre-focused elements from the focused elements and to signal the coming focus. It is also likely that speakers tried to save time and energy to produce the focused syllables. More studies should be carried out to investigate the pre-focused items and test these hypotheses.

## 5 Conclusion

The present study examined the effect of prosodic focus on stops in Chongming Chinese. In the on-focus condition, the phonological contrast between voiced and aspirated stops and unaspirated and aspirated stops were enlarged to indicate their lexical contrast. Post-focus influence was found in the VOT of voiced and unaspirated stops, suggesting that the domain of the VP focus may contain not only the verb but other arguments. Pre-focus adjustments of VOT were also found, which is suspected as preparations for the following focus. Further investigation is needed. Our study verifies that the influence of prosodic focus demonstrates cross-linguistic differences. Aspirated stops remain relatively stable in different focus conditions, revealing their use in Chongming Chinese is in lower frequency. Voiced stops demonstrated the feature of breathy phonation, which may explain its significant manipulation in focus conditions.

As the goal of the current study is to demonstrate some acoustic cues for differentiating different focus conditions, further perception study should be carried out to examine the link between perception and production and testify whether the differences of VOT in different focus conditions can actually be used as acoustic cues in human perception. It is suggested from other studies that other acoustic cues such as F0, also played a supplementary role in differentiating phonological categories. Future study should also consider other acoustic cues and compare the results with the current study.

In addition, due to the language-specific feature in both the VOT and the realization of prosodic focus, more languages should be involved in study and contribute new findings. More attention should be paid on the pre-focused region of utterances and seek more convincing explanation.

## Acknowledgments

This study was funded by a research grant from Faculty of Humanities, the Hong Kong Polytechnic University (grant number: 1-ZVHJ). We would like to thank all the informants and the three anonymous reviewers.

## References

Alzaidi M. Swaileh, Yi Xu and Anqi Xu. 2019. Prosodic encoding of focus in Hijazi Arabic. Speech Communication, 106, 127-149.

Boersma Paul. 2002. Praat, a system for doing phonetics by computer. Glot international, 5.

Chen Si. 2014. A Phonetic and Phonological Investigation of the Tone System of Chongming Chinese. [electronic resource]. University of Florida. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=cat04364a&AN=ufl.033650748&site=eds-live

Chen Szu-wei, Bei Wang and Yi Xu. 2009. Closely related languages, different ways of realizing focus. In Tenth Annual Conference of the International Speech Communication Association.

Chen Zhongmin. 2014. On the relationship between tones and initials of the dialects in the Shanghai area. In proceedings of 4th International Symposium on Tonal Aspects of Languages 2014,116-119.

Chen Yiya. 2011. How does phonology guide phonetics in segment– f0 interaction? Journal of Phonetics, 39(4), 612-625.

Cho Taehong and Patricia A. Keating. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. Journal of phonetics, 29(2), 155-190.

Cho Taehong and James M. McQueen. 2005. Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. Journal of Phonetics, 33(2), 121-157.

Choi Hansook. 2003. Prosody induced acoustic variation in English stop consonants. In Proceedings of the 15th International Conference of Phonetic Sciences 2003, 261-264.

Davenport Mike and Stephen J. Hannahs. 2013. Introducing phonetics and phonology. Routledge.

De Jong Kenneth. 2004. Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. Journal of Phonetics, 32(4), 493-516.

Hwang H. Kyung. 2012. Asymmetries between production, perception and comprehension of focus types in Japanese. In Proceedings of the 6th International Conference on Speech Prosody 2012, 614-644.

Jun Sun-Ah. 2011. Prosodic markings of complex NP focus, syntax, and the pre-/post-focus string. In Proceedings of the 28th West Coast Conference on Formal Linguistics (pp. 214-230). Somerville, MA: Cascadilla Press.

Jun Sun-Ah, Hee-Sun Kim, Hyuck-Joon Lee and Jong-Bok Kim. 2006. An experimental study on the effect

of argument structure on VP focus. Korean Linguistics, 13(1), 89-113.
Kuznetsova Alexandra, Per B. Brockhoff and Rune Haubo Bojesen Christensen. 2017. "lmerTest Package: Tests in Linear Mixed Effects Models." Journal of Statistical Software, 82(13), 1–26.
Lee Albert and Yi Xu. 2018. Conditional realisation of post-focus compression in Japanese. In Proceedings of the 9th International Conference on Speech Prosody 2018, 216-219.
Lee Yong-cheol. 2015. Prosodic Focus within and across Languages, (Doctoral dissertation). Available from ProQuest Dissertations and Theses database.
Ling Bijun and Jie Liang. 2016. The influence of syllable structure and prosodic strengthening on consonant production in Shanghai Chinese. In: 2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP), 1-5.
Lisker Leigh and Arthur S. Abramson. 1964. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. WORD, 20(3), 384-422.
Port Robert F. and Rosemarie Rotunno. 1979. Relation between voice-onset time and vowel duration. Journal of the Acoustical Society of America, 66(3), 654-662.
R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.URL http://www.R-project.org/.
Schneider Walter, Amy Eschman and Anthony Zuccolotto. 2002. E-Prime user's guide. Pittsburgh: Psychology Software Tools.
Wickham Hadley. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, http://ggplot2.org.
Xu Yi. 2011. Post-focus Compression: Cross-linguistic Distribution and Historical Origin. In Proceedings of the 7th International Conference on Phonetics Sciences 2018, 152-155.
Xu Yi. 2013. ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. In Tools and Resources for the Analysis of Speech Prosody (pp.7 – 10). Aix-en- Provence, France: Laboratoire Parole et Langage.
Xu Yi, Szu-Wei Chen and Bei Wang. 2012. Prosodic focus with and without post-focus compression: A typological divide within the same language family? The Linguistic Review, 29(1), 131-147.
Yang Yike, Si Chen and Kechun Li. 2018. Pitch realization of post-focus components in Chongming Chinese. The Journal of the Acoustical Society of America, 144(3): 1938-1938.
Yang Yike, Si Chen and Kechun Li. 2019. Effects of Focus on Duration and Intensity in Chongming Chinese. In Proceedings of ICPhS 2019, Melbourne.
Zhang Huiying. 1993 Chongming Fangyan Cidian [electronic resource]. Retrieved from http://img.chinamaxx.net.easyaccess2.lib.cuhk.edu.hk/n/abroad/hwbook/chinamaxx/