Prosodic Focus in Hong Kong Cantonese without Post-focus Compression

Wing Li Wu, Yi Xu

Department of Speech, Hearing and Phonetic Sciences, University College London, UK

wl.wu@ucl.ac.uk, yi.xu@ucl.ac.uk

Abstract

A recent study reported that post-focus compression (PFC) previously found in Beijing Mandarin is absent in two related languages, Taiwanese and Taiwan Mandarin, and that PFC is beneficial to focus recognition. This paper presents the results of acoustic and perception experiments for Hong Kong Cantonese, another Chinese dialect, which show that PFC does not occur in Hong Kong Cantonese, making it similar to Taiwanese and Taiwan Mandarin. Duration and intensity are found to be the two main acoustic correlates of prosodic focus in Hong Kong Cantonese, with pitch excursion size significant only in the dynamic tones. Focus recognition rate in Hong Kong Cantonese is high compared to the three languages above, and this suggests that other factors which were not examined, including the effect of prosodic structure on initial consonants, may also be important for the realization and recognition of prosodic focus.

Index Terms: Cantonese, focus, post-focus compression, focus perception

1. Introduction

Linguistically, a whole array of features can be employed by speakers for the purpose of making an emphasis in an utterance, from lexical choice to prosodic means. Prosodic focus can be achieved by varying the fundamental frequency (F_0), duration and intensity of the syllables in focus. It has been shown that F_0 variations are a major acoustic correlate of prosodic focus in non-tone languages [1, 2]. In the case of tone languages, F_0 variations are used for lexical distinction and so there is the likelihood that they are not used for encoding focus. But it has been demonstrated that such F_0 variations are an important means in the realization of focus in Beijing Mandarin [3, 4, 5, 6]. In Beijing Mandarin, there is a compression of the post-focus pitch range and intensity besides the expansion of the two acoustic features of the on-focus words. However, such post-focus compression (henceforth PFC) is not observed in Taiwan Mandarin [6]. Such finding reveals that even closely related tonal languages can have different realizations of focus. Since Taiwan Mandarin is found to be similar to Taiwanese in having no PFC [6], it is worth investigating whether other tone languages, in particular other Chinese dialects, show the same feature. Such experiments will have bearings on the typological classification of languages.

This paper presents the data of a production experiment and a perception experiment for Hong Kong Cantonese. Similar production experiments have been done which revealed that duration is the most significant acoustic correlate in Hong Kong Cantonese [7]. The present study performs a fuller analysis of all the lexical tones in Hong Kong Cantonese, and looks at whether PFC is present in the language, and compares the focus recognition rate with those reported for Beijing Mandarin, Taiwan Mandarin and Taiwanese [6].

Hong Kong Cantonese contains six lexical tones [8], which are commonly labeled Tones 1 to 6. Tones 2 and 5 are dynamic, both rising, and Tones 1, 3, 6 are static. Tone 4 is a low falling tone and can be treated as a static tone. If we use a 5-point scale to represent the tone contours, where 1 represents the lowest relative pitch level and 5 the highest, the six Hong Kong Cantonese lexical tones can be transcribed as 55, 25, 33, 21, 23 and 22.

2. Method

2.1. Production experiment

2.1.1. Stimuli

Six target declarative sentences were used, each consisting of three words composed of syllables having the same lexical tones, one target sentence for each lexical tone.

Table 1. *Target sentences*

Tones	Test sentences
1 (high level)	[ma˥ ma˥ mɔ˥ mau˥ mi˥] Mother touches the cat.
2 (high rising)	[siu˧˥ tsɛ˧˥ hou˧˥ tsou˧˥ tsɐu˧˥ tsɔ˧˥] The young lady has left very early.
3 (mid level)	[a˧ tʰai˧ tsɔi˧ tsʰi˧ him˧ tsai˧] Tai is in debt again.
4 (low falling)	[jɐn˨˩ kʷʰɐn˨˩ wɔ˨˩ pʰɪŋ˨˩ jɐu˨˩ hɐŋ˨˩] The crowd marched peacefully.
5 (low rising)	[lou˨˧ lei˨˧ man˨˧ man˨˧ mai˨˧ hai˨˧] Mr. Li buys crabs every night.
6 (low-mid level)	[hɔk˨ hau˨ tsʊŋ˨ si˨ wɐn˨ tʊŋ˨] The school values sports.

For the five-syllable Tone 1 sentence, the first two syllables form a word, the third syllable forms a word on its own, and the last two syllables form a word. In the other five sentences, every two syllables form a word. In the experiment, all these three-word sentences were assigned four different focus locations: no focus, initial (the first word), medial (the second word), and final (the last word), which were elicited by precursor questions each asking for a specific piece of information related to the target sentence.

2.1.2. Recording procedure and data extraction

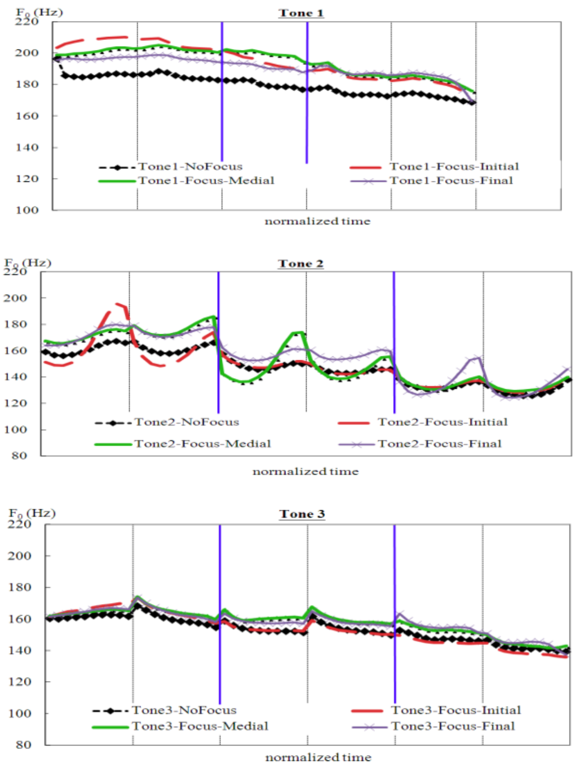
The recording was done in an anechoic room at University College London, UK. Eight native speakers of Hong Kong Cantonese (four males and four females aged between 19 and 33) took part as subjects. The test sentences and their precursor questions, generated from a randomized list, were shown one pair at a time on a computer screen. Each subject recorded each pair of precursor and test sentence five times in separate randomized blocks. The speech was captured using a sound level meter (Bruel and Kjaer 2231) as the microphone and recorded directly onto a computer hard disk with a sampling rate of 44.1 kHz.

Data extraction was done using a custom-written script [9] for the software Praat [10]. Praat provided automatic vocal pulse marking which was manually rectified using the script. For each test sentence, the script generated a minimally smoothed F_0 contour, and computed the mean F_0, mean intensity and duration for each syllable.

2.1.3. *Results and analyses*

Time-normalized mean pitch contours for all the six sentences are plotted in Figure 1. Except Tone 1, the static tones (Tones 3, 4 and 6) do not show much difference in pitch contours between on-focus words and neutral-focus words. The pitch contours for Tone 1 show a large increase in the pitch level on the focused words, more pronounced when the first word was emphasized. In fact, the pitch level of the whole sentence is shifted upwards in all the three sentences with focus. For the dynamic tones, it can be observed that the pitch contours of the on-focus words for Tone 2 are quite different from those in the neutral-focus sentences, with larger pitch range (excursion size) and higher pitch level in on-focus words, but Tone 5 does not show as obvious differences. In all six tones, no post-focus lowering in F_0 can be observed in any sentences with initial or medial focus. In general the graphs show that basically the on-focus words exhibit pitch variations but little focus-related variability is observed in the preceding or following words.

In order to verify which of the four acoustic features, namely mean F_0, F_0 excursion size, duration and intensity, are correlates of focus in Hong Kong Cantonese, and to confirm that there is no post-focus lowering of pitch range or intensity, repeated measures ANOVAs were carried out comparing the acoustic measurements at the same word location (initial, medial, final) across the four focus conditions (neutral, on-focus, pre-focus, post-focus) that the word can have (the initial word can be neutral, on-focus, and pre-focus; the medial word can be neutral, on-focus, pre-focus or post-focus; the final word can be neutral, on-focus or post-focus). Table 2 summarizes the ANOVA results. Post hoc Bonferroni tests were performed to pinpoint which of the following pairs are significantly different with respect to all four acoustic measurements: (1) neutral vs. on-focus, (2) on-focus vs. pre-focus, (3) on-focus vs. post-focus, (4) neutral vs. pre-focus, (5) neutral vs. post-focus, (6) pre-focus vs. post-focus.



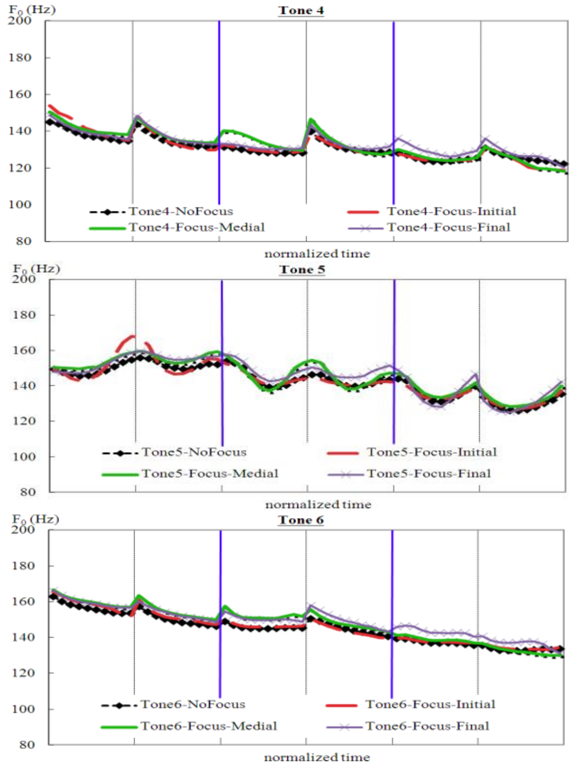


Figure 1. *Time-normalized mean F_0 contours of the six sentences, each curve representing an average of 40 repetitions by 8 subjects. The vertical lines show syllable boundaries, with the thicker ones indicating word boundaries.*

Table 2. *Repeated measures ANOVA results. The degrees of freedom (between and error) for both the initial and final locations are 2 and 14, and those for the medial location are 3 and 21. (abbreviations: loc. – word location; excur. – excursion size; I – initial word, M – medial word, F – final word; n.s.- not significant; ↑ – on-focus being higher than neutral focus shown in post hoc tests)*

tone & loc.	mean F_0	F_0 excur.	duration	intensity
	F *p*	F *p*	F *p*	F *p*
1 I	n.s.	3.73; ***0.050***	9.09; ***0.003*** ↑	27.0; ***<0.001*** ↑
M	n.s.	4.09; ***0.020***	13.1; ***<0.001*** ↑	39.5; ***<0.001*** ↑
F	n.s.	n.s.	17.6; ***<0.001*** ↑	10.9; ***0.001*** ↑
2 I	5.63;***0.016***↑	51.8; ***<0.001*** ↑	33.2; ***<0.001*** ↑	16.6; ***<0.001*** ↑
M	5.32;***0.007***	20.3; ***<0.001*** ↑	26.2; ***<0.001*** ↑	13.9; ***<0.001*** ↑
F	n.s.	28.1; ***<0.001*** ↑	17.1; ***<0.001*** ↑	27.4; ***<0.001*** ↑
3 I	n.s.	n.s.	34.7; ***<0.001*** ↑	21.6; ***<0.001*** ↑
M	6.67;***0.002***↑	3.89; ***0.024*** ↑	25.9; ***<0.001*** ↑	19.0; ***<0.001*** ↑
F	4.92;***0.024***	10.6; ***0.002*** ↑	16.7; ***<0.001*** ↑	25.1; ***<0.001*** ↑
4 I	n.s.	5.86; ***0.014***	33.9; ***<0.001*** ↑	19.0; ***<0.001*** ↑
M	3.15;***0.046***↑	12.1; ***<0.001*** ↑	21.0; ***<0.001*** ↑	n.s.
F	8.28;***0.004***↑	5.48; ***0.017*** ↑	12.0; ***0.001*** ↑	38.3; ***<0.001*** ↑
5 I	n.s.	26.0; ***<0.001*** ↑	31.3; ***<0.001*** ↑	17.5; ***<0.001*** ↑
M	3.58;***0.031***	52.0; ***<0.001*** ↑	19.6; ***<0.001*** ↑	12.0; ***<0.001*** ↑
F	n.s.	30.4; ***<0.001*** ↑	16.9; ***<0.001*** ↑	19.4; ***<0.001*** ↑
6 I	n.s.	6.12; ***0.012*** ↑	67.4; ***<0.001*** ↑	17.3; ***<0.001*** ↑
	3.66;***0.029***↑	6.07; ***0.004***	42.2; ***<0.001*** ↑	14.4; ***<0.001*** ↑
	12.3;***0.001***↑	9.25; ***0.003*** ↑	19.5; ***<0.001*** ↑	26.8; ***<0.001*** ↑

Table 2 shows that both duration and intensity exhibit significant differences across the four focus conditions in all six tones, in all three word locations. F_0 excursion size displays significant differences in at least two word locations in each of the six tones, whereas for mean F_0, the Tone 1 sentence does not show significant differences, but significant differences are found for other tones in at least one word location. Some of the significant results of the repeated measures ANOVAs shown in Table 2 turned out to be not significant in pairwise comparisons, and this may be due to the use of a fairly conservative post hoc test.

In terms of the post hoc tests comparing on-focus and neutral conditions, there are significant differences in both duration and intensity, both greater in magnitude for the on-focus words. This is the case for all six tones and all three word locations. (In addition, the on-focus words have significantly longer duration than pre-focus and post-focus words in all instances, and greater intensity in nearly all instances). In contrast, only three instances of significant differences are observed for mean F_0 (Tone 3, medial word; Tone 4, final word; Tone 6, final word; the mean F_0 of on-focus is higher than that of neutral focus in all these instances). For F_0 excursion size, instances of significance are found in all tones except Tone 1 sentences. Specifically, excursion size of all three word locations in Tones 2 and 5 (i.e., the dynamic tones) is found to be significantly different between the on-focus words and neutral focus words (the other significant instances include: Tone 3, final word; Tone 4, medial and final words; Tone 6, final word). In all these significant cases, the F_0 excursion size is higher in on-focus words than in neutral focus words.

As to whether there is post-focus lowering of mean F_0 and intensity, the post hoc comparisons between post-focus and neutral conditions will provide the answer. It is found that for mean F_0, there is significant difference in only one instance (Tone 6, medial word; the post-focus word having a higher mean F_0 than neutral focus words), whereas there is no significant difference in intensity. That is to say, no lowering in either mean F_0 or intensity can be said to occur post-focally in Cantonese.

It is therefore clear from the statistical analyses that both duration and intensity are the major acoustic correlates of focus in the language. Table 3 presents the means of syllabic duration and intensity of the neutral-focus words and on-focus words across the six tones.

Table 3. *The means of duration and intensity (per syllable) of the neutral-focus words and on-focus words (same abbreviations as in Table 2)*

tone & loc.	duration (milliseconds)		intensity (decibels)	
	neutral	on-focus	neutral	on-focus
1 I	204.4	244.8	72.5	75.6
M	238.1	340.8	71.3	75.3
F	251.4	313.0	69.7	72.6
2 I	124.2	196.6	70.5	72.9
M	125.9	194.6	68.2	71.1
F	171.8	242.3	65.1	69.7
3 I	122.1	185.8	71.0	75.0
M	126.0	188.3	68.9	72.7
F	207.3	265.8	66.9	71.1
4 I	176.1	242.9	67.4	70.8
M	162.4	223.5	65.8	69.9
F	181.4	233.5	63.2	68.5
5 I	188.7	264.4	70.2	73.2
M	224.2	303.2	67.9	71.5
F	240.6	307.2	65.0	68.4
6 I	107.1	153.3	70.8	75.2
M	119.4	190.0	67.7	70.9
F	193.7	272.3	65.9	70.5

Mean F_0 variation does not play a significant role in signaling focus. Although Figure 1 shows that in the Tone 1 sentences, the F_0 level is raised in the sentences with focus, the statistical tests do not show significant differences; the F_0 plots of individual speakers reveal that there are indeed much individual variation in using an overall raised pitch for emphasizing Tone 1 words. Excursion size is significant in the dynamic tones in all three word locations, and also in the level tones, except the high level tone (Tone 1), in some word locations. Since duration is increased in all focused words as shown above, the longer duration probably has enabled the dynamic tones to have a much larger increase in F_0 excursion size than the static tones [14], and this is reflected by the significant differences seen in all word locations in both dynamic tones.

2.2. Perception experiment

2.2.1. *Stimuli, subjects and listening procedure*

The stimuli used in the perception experiment were sentences taken from the production experiment. The recordings of the three speakers who showed maximum, median and minimum standard deviations of all F_0 points across all the focus conditions by all speakers were selected. Consequently, there were a total of 360 tokens in the listening test (3 speakers, 5 repetitions, 4 foci, 6 sets of sentences).

16 native speakers of Hong Kong Cantonese, recruited at the Hong Kong Institute of Education, participated in the experiment. The experiment was carried out in a quiet room using the MFC function of Praat. Each subject wore a pair of headphones and listened to the randomized test sentences and decided on which one or none of the three words in each test sentence was emphasized. The confusion matrices are shown in Table 4.

2.2.2. *Results*

Table 4. *Confusion matrices of focus perception (%). The correct identifications are indicated by boldface.*

sentence	original \ heard as	None	Initial	Medial	Final
Tone 1 (high level)	None	**77.92**	12.08	6.25	3.75
	Initial	34.17	**57.92**	7.08	0.83
	Medial	35.00	5.42	**58.33**	1.25
	Final	27.50	2.08	18.33	**52.08**
Tone 2 (high rising)	None	**84.17**	8.33	3.75	3.75
	Initial	20.42	**73.33**	3.75	2.50
	Medial	17.92	2.08	**78.75**	1.25
	Final	9.58	0.83	2.08	**87.50**
Tone 3 (mid-level)	None	**81.67**	7.50	7.50	3.33
	Initial	20.83	**69.58**	8.33	1.25
	Medial	7.08	2.92	**89.58**	0.42
	Final	12.92	0.42	2.08	**84.58**
Tone 4 (low falling)	None	**86.25**	7.50	5.00	1.25
	Initial	28.33	**69.58**	2.08	0.83
	Medial	13.33	1.67	**84.58**	0.42
	Final	13.33	0.83	6.67	**79.17**
Tone 5 (low rising)	None	**82.50**	8.33	5.00	4.17
	Initial	10.83	**87.08**	2.08	0.00
	Medial	11.25	1.67	**86.25**	0.83
	Final	18.33	0.83	3.75	**77.08**
Tone 6 (low-mid level)	None	**81.25**	5.00	9.17	4.58
	Initial	33.33	**57.50**	6.67	2.50
	Medial	16.67	1.67	**80.83**	0.83
	Final	9.58	0.83	2.50	**87.08**

Table 3 shows that the overall focus identification rate is fairly high, except for Tone 1, for which the average rate is less than 60%. In addition, the identification rate of initial focus was generally poorer than that of medial or final focus, which is different from the pattern seen in Beijing Mandarin [6].

When focus was heard wrongly, in most cases the listeners heard no focus in the sentence. In cases when there was no focus in the sentence but the listeners thought that there was one, any of the three locations (initial, medial, final) could be perceived as having the focus, but in five out of the six tones the perceived focus was in the initial position.

3. Discussion

The acoustic difference between on-focus and neutral focus in Cantonese is clear: duration and intensity are the main acoustic correlates of prosodic focus and both are increased significantly in the on-focus words in any word location for all lexical tones. Besides, no decrease in mean F_0 or intensity is found in the post-focus words. Therefore, Hong Kong Cantonese is similar to Taiwanese and Taiwan Mandarin with respect to how focus is realized phonetically. Hong Kong Cantonese speakers do not seem to manipulate fundamental frequency as a means to convey focus, especially for the static tones (Tones 1, 3, 4 and 6), but may expand the pitch range for the two dynamic tones (Tones 2 and 5), in which case it is likely to be a secondary effect of increased duration. The data suggest that speakers of Hong Kong Cantonese tend not to employ the F_0 of the lexical tones for the purpose of emphasis, but keep them relatively unchanged as in non-focus conditions. This finding is consistent with previous reports that there is minimal effect of focus on the tones in Cantonese [11, 12].

Chen et al. [6] recently showed that with the compression in both pitch range and intensity of post-focus words in Beijing Mandarin, higher focus recognition rates were achieved by Beijing Mandarin speakers than speakers of Taiwanese and Taiwan Mandarin (recognition rate of >90% versus <75%). Their test sentences consist of five syllables all with the high level tone and essentially the same words as the Tone 1 test sentence used in this paper. The present study on Cantonese shows that focus identification was poor for the high level tone, with a recognition rate of less than 60% in all three word locations. However, the high focus recognition rates of approaching 70-80% in most other tones suggest that there may be other factors besides PFC as effective cues in focus perception.

Based on the data presented in this paper, we can consider several speculations for explaining the poor performance of Tone 1 focus identification. Firstly, the acoustic analysis reveals that F_0 excursion of on-focus words is significantly different from that of words in neutral focus in all tones except Tone 1, and so this distinction between Tone 1 and other tones can explain the poorer focus recognition rate. Secondly, increases in both duration and intensity are the most prominent feature of focus in Cantonese, and if intensity differences are indeed important for focus recognition, then Tone 1 syllables would be the least advantageous, because a higher F_0 is associated with a higher level of intensity, and the already high F_0 in Tone 1 syllables would exhibit relatively smaller percentage increase in intensity when they are on-focus than syllables of other lexical tones. Thirdly, the Tone 1 syllables all begin with the sonorant consonant [m], whereas in the other five test sentences, there is a mixture of sonorant and obstruent consonants in syllable initial positions. This difference may explain why Tone 1 sentences had the poorest focus recognition rate, i.e., it could be due to a lack of consonantal cue present in other test sentences. Since acoustic properties such as voice onset time, stop burst energy and consonantal length may act as cues for focus perception [13], they might have contributed to the higher recognition rates of focus in sentences of Tones 2 to 5 in the experiment, and this will require a further experiment to confirm. Regardless of whether consonantal effects are significant, the direct comparison of the Cantonese data with the data of Beijing Mandarin, Taiwan Mandarin and Taiwanese suggest the importance of PFC, at least for words with the high level tone.

4. Conclusions

The production experiment demonstrated that duration and intensity are the major acoustic correlates of prosodic focus in Hong Kong Cantonese; F_0 is not significantly modified for signaling focus, and F_0 excursion is increased in the dynamic tones but not in the static tones. Results of the listening test, when compared with the data on Beijing Mandarin, provided clues that post-focus compression of F_0 and intensity could be important acoustic cues for focus recognition but also points to the possibility that consonantal features also play an important part in enhancing focus perception accuracy.

5. References

[1] Rump, H.H., and Collier, R. (1996). “Focus conditions and prominence of pitch-accented syllables,” *Language and Speech*, 29, 1-17.

[2] Botinis, A., Fourakis, M., and Gawronska, B. (1999). “Focus identification in English, Greek and Swedish,” in *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, USA, vol. 2, 1557-1560.

[3] Jin, S. (1996). *An acoustic study of sentence stress in Mandarin Chinese*. Ph.D. dissertation, The Ohio State University.

[4] Xu, Y. (1999). “Effects of tone and focus on the formation and alignment of F_0 contours,” *Journal of Phonetics*, 27, 55-105.

[5] Liu, F., and Xu, Y. (2005). “Parallel encoding of focus and interrogative meaning in Mandarin intonation,” *Phonetica*, 62, 70-87.

[6] Chen, S-W, Wang, B., and Xu, Y. (2009). “Closely related languages, different ways of realizing focus”, in *Proceedings of the 10th Annual Conference of the International Speech Communication Association* (Interspeech 2009), Brighton, UK, 1007-1010.

[7] Bauer, R.S., Cheung, K.H., Cheung, P.M., and Ng, L. (2004). “Acoustic correlates of focus-stress in Hong Kong Cantonese.” In Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistics Society 2001, Tempe, Arizona State University Program for Southeast Asian Studies Monograph Series, pp. 29-49.

[8] Zee, E. (1999). “Chinese (Hong Kong Cantonese)”, in The International Phonetic Association, *Handbook of the International Phonetic Association*, 58–60. Cambridge, U.K.: Cambridge University Press.

[9] Xu, Y. (2005-2009). “_TimeNormalizeF0.praat. Online: http://www.phon.ucl.ac.uk/home/yi/tools.html”

[10] Boersma, P. (2001). “Praat, a system for doing phonetics by computer,” *Glot International*, 5 (9/10), 341-345.

[11] Man, V. C. H. (2002). “Focus effects on Cantonese tones: An acoustic study.” in *Proceedings of the first international conference on speech prosody*, Aix-en-Provence, France, 467–470.

[12] Gu, W. and Lee, T. (2007). “Effects of focus on prosody of Cantonese speech – A comparison of surface feature analysis and model-based analysis.” in *Proceedings of the International Workshop Paralinguistic Speech ’07*, Saarbrücken, Germany, 59-64.

[13] Cho, T. and Keating, T.A. (2001). “Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, 29, 155-190.

[14] Xu, Y. and Wang, M. (2009). “Organizing syllables into groups – Evidence from F0 and duration patterns in Mandarin. *Journal of Phonetics*, 37, 502-520.