Quantitative analysis of preboundary lengthening in Cantonese

Ivan Chow

Theoretical and Applied Linguistics Laboratory
The University of Western Ontario
ichow3@uwo.ca

Abstract

In a corpus study using eight pairs of syntactically ambiguous sentences, we conducted acoustic and quantitative analyses of the prosodic patterns used by seven native speakers of the Hong Kong variety of Cantonese to divide these sentences into prosodic words — a phenomenon referred to as "prosodic chunking". Although pauses, pitch reset and preboundary lengthening were analyzed, in this article we concentrate on presenting results from the analysis of preboundary lengthening. Using test sentence pairs consisting of identical series of words with two possible prosodic subdivisions, we measure the preboundary syllables within the sentence, and compare it to the same syllables in the non-boundary position of the corresponding sentence. Results indicate that the presence/absence of a following prosodic boundary is highly significant in the measure of lengthening, thus confirming preboundary lengthening as a prominent device in marking prosodic word boundaries in Cantonese. Moreover, the presence of a pause at the prosodic word boundary also triggers a more prominent preboundary lengthening. Finally, our statistical results indicate that there seems to be a trade-off relation between pitch reset and preboundary lengthening. Since this result seems to contradict with recent research, which indicates that pitch range *increases* with syllabic duration [1], and that pitch contours are also subject to contextual tonal effects (both carry-over and anticipatory) [2] as well as perturbations brought about by focus [3]. As a result, more research is needed before we can confirm/disconfirm the validity of this trade-off relation.

1. Introduction

"Prosodic chunking" refers to the phenomenon where speakers use different prosodic devices (e.g. preboundary lengthening, pitch reset [4], pauses) to divide an utterance into prosodic units (or chunks) (cf. [5]). The study of prosodic chunking is of interest to both theoretical linguistics and the development of speech technology; in understanding the interface between phonetic forms and meaning; and in improving the naturalness and effectiveness of speech synthesis and recognition. Pre-boundary lengthening has been reported as a prosodic juncture marker in different languages ([6] for Dutch; [7] for French; [8] and [9] for Mandarin). Many studies have been carried out to examine prosodic chunking behaviours in Mandarin. For example, in a study using syntactically ambiguous test sentences, Shen [9] finds preboundary lengthening to be a prominent boundary marker between prosodic phrases. Using six-syllable test sentences that can be divided into two 3-syllable prosodic words or three 2-syllable prosodic words, Chow [7],[8] finds both pitch reset and preboundary lengthening to be prominent markers of prosodic boundaries between prosodic words. Chiang et al. [10] maintain that different types of prosodic boundary markers (e.g. pitch reset, preboundary lengthening and pauses) are present at boundaries of different strengths; the amplitude of prosodic boundary markers varies depending on the strength of the boundaries.

Although both Mandarin and Cantonese are Chinese languages and present numerous similarities in pronunciation, the complexity of Cantonese lexical tones (both melodically and rhythmically; see section 2 for a detailed discussion of the Cantonese lexical tone system) makes the study of prosodic chunking in Cantonese especially challenging. While many studies on Mandarin prosodic chunking have been published (e.g. [8], [9], [10], etc.), only a few have been conducted on Cantonese prosodic chunking at this point. In a pilot study, Chow [11] conducted an acoustic analysis of the recordings from three native speakers of the Hong Kong variety of Cantonese. Preboundary lengthening was found to be a prominent marker of prosodic word boundaries. However, the frequency and amplitude of preboundary lengthening seem to vary significantly between speakers, which points to the fact that a larger-scale study is necessary in order to improve the statistical credibility of these findings. Hence a much larger speech corpus is created for the present study. We analyze the recordings of seven native speakers (3M4F) of the Hong Kong variety of Cantonese reading eight pairs of specially designed syntactically ambiguous test sentences (see section 3 for a detailed discussion of the test sentence design). A total of 660 tokens are analysed. For each token, we measure the effects of pitch reset, preboundary lengthening and pauses. However, in this article we concentrate on presenting results from our analysis of preboundary lengthening.

In section 2, we give a brief introduction to the Cantonese lexical tone system; discuss the differences between Mandarin and Cantonese tones and syllabic structures as well as their differences in the prosodic patterns. A discussion of the test sentence design is found in section 3. The experimental methodology is outlined in section 4, with a discussion of the test results in section 5 and the conclusions given in section 6.

2. Cantonese lexical tones

As shown in table 1, there are six lexical tones in the Hong Kong variety of Cantonese. A morpheme in Cantonese consists of the combination of a syllable and a lexical tone. It functions as both a phonological and morphological unit of analysis [12]. With the exception of a small group of functional morphemes, morphemes are formed by combining a syllable with one of the six tones. The meaning of the morpheme is changed when a given syllable is combined with different tones. (The Jyutping tone system (cf. [1-3]) is used in the tone transcription in this article. Six citations tones are used in this system (as seem in figure 1), while the 3 addition 'checked' tones are treated as shorter versions of three level tones, namely tones 1, 3 and 6.)

High Tones	**(1) High Level**	**(2) High Rising**	**(3) Mid Level**
	yan1 因 (reason)	yan2 忍 (to endure)	yan3 印 (a stamp)
Low Tones	**(4) Low falling (level)**	**(5) Low Rising**	**(6) Mid-Low Level**
	yan4 人 (people)	yan5 引(to lead)	yan6 孕 (pregnancy)

Table 1. Cantonese lexical tones (non-checked)

While the six tones are associated with open syllables and syllables ending with nasal consonants (/n/, /m/ and /ŋ/), there are also 'checked syllables' in Cantonese. These are syllables ending with unreleased stops (/p/, /t/ and /k/), and are inherently shorter in duration [13], [14], [15]. Checked syllables are combined with the mid-low level, mid-level and high level tones only. As illustrated in figure 2, the high level (1) and mid level (3) tones are mutually exclusive. In other words, within the 3-tone repertoire, a given checked syllable can only combine with the mid-low level, and either the mid-level or the high-level tones.

High Level Tone	sap1 濕 (wet)	**yeuk1 ***	yat1 一 (one)
Mid Level Tone	**sap3 ***	yeuk3 約 (contracts)	**yat3 ***
Mid-Low Level Tone	sap6 十(ten)	yeuk6 弱 (weak)	yat6 日(the sun, day)

Table 2. Cantonese checked tones

Now we turn to discussing the differences between the Mandarin and Cantonese lexical tones systems. First of all, there are four lexical tones in Mandarin: high-level, mid-rising, falling-rising, high-falling. These tones can be distinguished by tone shape alone. Whereas there are 6 lexical tones in Cantonese, which are distinguished both by tone shape and relative pitch height [12], [13], [14] — there are two rising tones (2 and 5) and four level tones (1, 3 and 6; tone 4 is often realized as a very low level tone as well), which can be distinguished only by relative pitch height. While Mandarin tones are distinguished by tone shape alone, both tone shape and relative pitch height are distinguishing features in Cantonese. Secondly, as the syllabic repertoire of Mandarin consists of open syllables and those ending with nasals or glides only, the Cantonese syllabic repertoire consists of open syllables, those ending with glides, nasals as well as non-released stops (checked syllables); checked syllables are inherently shorter. This makes syllabic duration a third distinguishing feature in this language. Since syllabic duration is involved in both lexical tone recognition at the word level, and in preboundary lengthening at the prosodic word and higher levels, an interesting question arises as to whether syllabic duration can serve as a distinguishing feature at both levels without creating miscommunications? Or does one of the two (preboundary lengthening or word-level melody movements) give way in favour of the other?

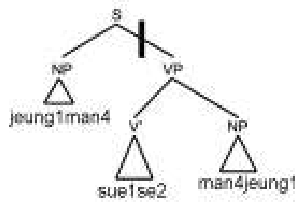
3. Design of the test sentences

Eight pairs of test sentences are used for this acoustic study. Within each sentence pair, corresponding sentences consist of an identical array of six morphemes. In the test sentences, the boundary between major syntactic constituents (e.g. the subject NP and the VP predicate) can be placed at two different locations. Moreover, a third boundary can be placed between the VP and its object NP in the sentence that is divided into three 2-syllable prosodic words. The meaning of the sentence is changed when the boundary location is altered. In each sentence pair, syntactic boundaries are placed either between syllables 2 and 3, 4 and 5 in one case), or between syllables 3 and 4 (as shown in figure 3). This way, prosodic changes can be analyzed by direct comparison—where a boundary is present after syllable 2 in structure (a), it is not present at the same location in structure (b). Melodic and durational variations at this particular location can then be measured and compared directly against the corresponding sentence to determine the way in which the syntactic boundary was prosodically conveyed. Using identical series of morphemes in both alternatives, we are then able to minimize contextual variations (in terms of pitch and syllabic duration) due to different morpheme and lexical tone combinations. Consequently, differences in prosodic patterns of the two corresponding sentences can be readily attributed to prosodic juncture marking.

(3a) Jeung1 man4 / sue1 se2 man1 jeung1.
張文/書寫文章。
"Jeung-man writes articles."

(3b) Jeung1 man4 sue1 / se2 man1 jeung1.
張文書/寫文章。
"Secretary Jeung writes articles."

(3a)



(3b)

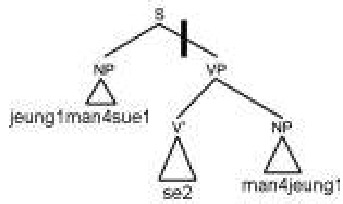


Figure 3. Sample of a pair of test sentences and their respective syntactic structures

In figure 3, (3a) and (3b) show one of the sentence pairs used in this experiment. These sentences are designed in such a way that the third syllable can be part of the VP-predicate as in structure (3a); or it can be part of the NP-subject as in structure (3b). Depending on the (syntactically-motivated) prosodic grouping, the sentence has exactly two interpretations; no other permutations of boundary location would render sensible interpretations other than the ones listed above.

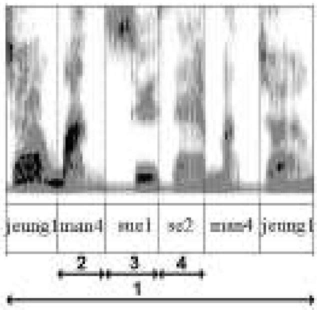
4. Methodology

For the phonetic experiment, subjects are asked to sit in a quite room and are given a printed script containing the test sentences. The two possible prosodic groupings are clearly marked on the script by slashes. The semantic differences of the test sentences and the purpose of the marked boundaries are explained to them. After a few practice runs, they are asked to read the script aloud as if they were engaged in a natural conversation while their voices are being recorded using a Sony digital minidisk recorder (MZ-N707). Each test sentence pair is repeated twice. Excluding speech errors, a total of 660 tokens are collected and analyzed. Acoustic analysis of the recorded speech is conducted using PRAAT [15] digitized at a sampling rate of 22 kHz for analysis.

4.1. Acoustic analyses

The recorded signals are analyzed using PRAAT. Borders are then placed between segments of difference sizes (e.g. prosodic word, syllable, phonemes, etc.). Time-tags are automatically generated by PRAAT, which are used for measuring syllabic duration.

(a).



(b).

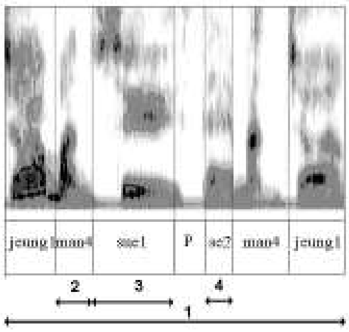


Figure 4. Measurement of syllabic duration

Figure 4 shows an example of a test sentence pair within which the preboundary syllables are measured and compared. Duration of preboundary syllables is measured in both (a) and (b), i.e. syllables 2 and 4 for sentence (a) where the prosodic boundaries are present with no pauses; syllable 3 for sentence (b) where "P" indicates the presence of a pause (boundary), These measurements are then divided by the duration of the entire sentence (1) to compensate for speech rate difference between corresponding sentences. Since the duration of pauses can vary tremendously between individuals, they are excluded from the duration of sentences—the duration of the sentence is calculated as the sum of the duration of all six syllables. A percentage value is then derived as a measurement of the proportion of the preboundary syllable within the sentence. This value is henceforth referred to as "percentage proportion". Subsequently, the percentage proportions are compared between the preboundary and non-preboundary syllables to determine whether preboundary lengthening is present at a boundary location. Since syllabic duration is a contrastive feature in Cantonese, the comparative method used here serves to minimize the impact of contextual variations in different morpheme combinations that would otherwise be present in a non-controlled environment.

For the purpose of the correlational test, a measure of the amplitude of preboundary lengthening (referred to as "comparative percentage proportion") is derived by calculating the difference between the percentage proportion of the preboundary syllable and that of the same syllable in the corresponding sentence. A positive value means a more prominent preboundary lengthening at the prosodic boundary. As a preliminary measure of the amplitude of pitch reset, we measure the difference in pitch value (F_0) at the optimal point of intensity at the time between the syllables across a prosodic boundary (referred to as "pitch range difference" cf. [4]). However, [1-3] indicate that as a lexical tone language, the contour of the tone (at the onset, mid-point and offset), tone-syllable alignment, the tonal context, the identity of the initial consonant of the following syllable, as well as focus all play a role in influencing the measurement of pitch reset. Knowing that the measure of pitch at the optimal-intensity point is confounded by different factors in play in Cantonese, we are conducting a detailed study on the effects of "prosodic chunking" on pitch reset.

4.2. Quantitative analyses

The encoded textgrid is then exported to Microsoft Excel to prepare spreadsheets for statistical analysis. The percentage proportion is treated as the dependent variable. Amongst other relevant factors, the presence/absence of prosodic word boundary, the presence/absence of pauses, the identity of the speaker as well as the identity of the trials are treated as independent variables. We then conduct quantitative analyses using SPSS. Four one-way ANOVA are carried out to determine the significance of the independent factors. We also conduct a correlation test between two independent variables: comparative percentage proportion of the preboundary syllable (a measure of the amplitude of preboundary lengthening), and comparative pitch range difference (a measure of the amplitude of pitch reset) in order to determine the correlation (if any) between the two prosodic boundary markers. Results are given in the next section.

5. Results and discussions

Independent variables	F	σ
Pres/Abs Boundary	66.772	0.000
Pres/Abs Pause	94.731	0.000
Speaker Identity	1.079	0.373
First/Second trial	0.149	0.699

Table 5. Summary of results from the quantitative analyses

Table 5 shows a summary of the results of the one-way ANOVA tests. The presence/absence of boundary is a highly significant factor. On average, preboundary syllables are lengthened by 2.34% (normalized using comparative percentage proportion; standard deviation = 4.43%) as compared to their counterparts in the non-boundary location ($\sigma = 0.00$). As a result, preboundary lengthening is a prominent prosodic word boundary marker in Cantonese. The strong significance of the presence/absence of a pause also indicates that the presence of a pause triggers a stronger preboundary lengthening. In all prosodic boundaries, preboundary syllables are lengthened on average 2.66% (standard deviation = 4.58%) when a pause is present, whereas these syllables are lengthened 1.52% (standard deviation = 3.94%). Speaker identity turns out to be an insignificant factor, which means that compared to Chow's pilot study [8] with 3 subjects and 24 tokens, the number of subjects (7) and the number of tokens collected (330) in this study are large enough to eliminate any individual differences found in the previous study. Consequently, the observations made in this study can be sufficiently generalized to all native speakers of the Hong Kong variety of Cantonese. Since no significant difference is found between the first and second trials, no learning effect is observed.

In the correlational test between comparative percentage proportion and comparative pitch range difference, a total of 330 tokens are analyzed. Results seem to indicate that there is a significant negative correlation between the two variables ($\sigma = 0.022$ (two-tailed); correlation = 1:-0.126) between the amplitude of preboundary lengthening and pitch reset; which indicates the presence of a trade-off effect. However, Wong [3] indicates that pitch values in Cantonese are subject to assimilatory carry-over effects, as well as a smaller but significant anticipatory effect of tonal dissimilation. In addition to contextual tonal effects, Gu & Lee [2] shows that in Cantonese, focus can affect pitch contours by raising the F_0 values and expanding the F_0 range. House & Fairbank [1] also indicate that the identity of the initial consonant of the following syllable can also trigger anticipatory effects to the pitch contour of the preceding syllable. As such, using pitch range difference at the optimal intensity point as a measure of pitch reset (as it is done in measuring pitch reset in non-tonal languages [4], [5]) seems to be confounded by many other factors present in the more complex Cantonese prosodic pattern. As a result, we need more detailed analyses of the pitch contour at the onset, mid-point and off-set of the syllable while taking into account different contextual tonal effects, before we are ready to confirm or disconfirm the validity of such a correlation between pitch reset and preboundary lengthening.

6. Conclusions

The study of prosodic chunking is of particular importance because it is a universal feature of human languages, and yet it is susceptible to influences of the idiosyncratic prosodic patterns of individual languages. Despite the fact that Cantonese makes use of tone shape, relative pitch and syllabic duration in word (tonal) recognition, results from this study show consistent behavioural patterns in the use of preboundary lengthening to those reported in studies on Mandarin [7], [8], [9], [10]. Results from the quantitative analyses indicate that preboundary lengthening is a prominent prosodic word boundary marker in Cantonese. The presence of a pause also triggers preboundary lengthening of larger amplitude. Finally, although a trade-off effect is observed between pitch reset and preboundary lengthening. More detailed analyses of the contextual pitch variations between neighbouring syllables (across boundaries) are needed before we can confirm/disconfirm the validity of such correlation.

In addition to preboundary lengthening, we are currently analyzing the effects of pitch reset at prosodic word boundaries. Our preliminary results indicate that the combination of lexical tones across the boundary is a significant factor in determining the amplitude of pitch reset. Given that Cantonese makes use of relative pitch height as a distinguishing feature in lexical tone recognition, it is not surprising that pitch reset is affected by lexical tones. Moreover, the presence of stops in checked tones also triggers micro-prosodic movements in the following syllable [16]. As such, we are working on teasing apart the effects of lexical tones in order to understand the interactions between word-level and higher-level prosodic patterns in Cantonese.

7. References

[1] House, A.S. & Fairbank, G. 1953. "The influence of consonant environment upon the secondary acoustical characteristics of vowels", *Journal of the Acoustical Society of America*, 25, 105-113.

[2] Gu, W.T. & Lee, T. 2007. "Effects of tonal context and focus on Cantonese F_0", *Proceedings of 16th International Congress of Phonetic Sciences* (ICPhS XVI), 1033-1036.

[3] Wong, Y.W. 2006. "Contextual tonal variations and pitch targets in Cantonese", *Proceedings: Speech Prosody 2006 (SP2006)*. Dresden, Germany. 317-320.

[4] Swerts, M. G. J., 1997. "Prosodic features at discourse boundaries of different strength", *Journal of the Acoustical Society of America*, 101(1): 514-521.

[5] Carlson, R.; Swerts M., 2003. Perceptually based prediction of upcoming prosodic breaks in spontaneous Swedish speech materials. *Proceedings 15th International Congress of Phonetic Sciences, Barcelona,* 507-510.

[6] Gussenhoven, C. & Rietveld, A. C. M., 1992. "Intonation contours, prosodic structure and preboundary lengthening", *Journal of Phonetics*, 20(3): 283-303.

[7] Chow, I., 2003. *Prosodic structure in French and Mandarin*, Ph.D. Thesis. University of Toronto.

[8] Chow, I., 2005. "Resolving temporary syntactic ambiguity by prosodic devices in Mandarin: An Acoustic Study", *Journal of Chinese Linguistics*, no. 33, vol. 2.

[9] Shen, X. S., 1992. "A Pilot study on the relation between the temporal and syntactic structure in Mandarin", *Journal of the International Phonetic Association*, 22(1/2): 35-43.

[10] Chiang, C-Y.; Yu, H-M.; Wang, Y-R.; Chen, S-H., 2007. "An Automatic Prosodic Labeling Method for Mandarin Speech". *Proceedings: Interspeech 2007*. 494-497.

[11] Chow, I., 2005. "Prosodic cues for Syntactically-Motivated Junctures". *Proceedings: Interspeech 2005*. 2373-2376.

[12] Chan, M. K. M., 1999. "Review of: Bauer, Robert S. and Paul K. Benedict (1997). Modern Cantonese Phonology. Berlin and New York: Mouton de Gruyter." Cahiers de Linguistique Asie Orientale, 28.1:101-112, The Ohio State University.

[13] Gandour, J., 1981. "Perceptual dimensions of tone: evidence from Cantonese", *Journal of Chinese Linguistics*, 9: 20-36.

[14] Mai, Y. 麥耘. 1998. "廣州話語調說略 [A brief introduction to Cantonese tones], 廣州話研究與教學 [Research and Pedagogy of Cantonese], 3:269-280.

[15] Boersma, P.; Weenink, D., 2007. *Praat: doing Phonetics by computer. Version 4.6.36.* [Computer Program]. Institute of Phonetic Sciences. University of Amsterdam.