Tone Spans of Cantonese English

Suki S.Y. Yiu

Linguistics Department, University of Hong Kong
syutji@hku.hk

Abstract

This paper examines the newly developed tones in Cantonese English, with a particular interest in the tones spanning across syllables. It attempts to provide phonetic evidence for the tone spans and demonstrate the association of tones to syllables.

Audio-recording data elicited from six speakers of Cantonese English was processed with Praat, and then fitted to a smoothing spline analysis of variance with R to generate smoothing splines at 95% confidence intervals for determining if the tones in different positions of a word are significantly different from each other. Pitch tracks of tones in individual words were also generated for a detailed examination of the tone contours of the tone spans.

Results showed that the realization of different tones was restricted by the position of the syllables in a word. Tone spans, Ω spans, H spans, and M spans were discussed and illustrated with tonal associations.

Index Terms: tone, tone spans, Cantonese English, Hong Kong English

1. Introduction

Cantonese English is a variety of language that owns its distinctiveness in many aspects of language which are worth paying attention to for the good of providing a neutral and unbiased description in its own terms [1]. In fact, many aspects of the phonology of Cantonese English have been studied so far [2-6]. Specifically, the use of pitch in Cantonese English, unlike tones in other languages, is perceptually distinct but carries a lower functional load, with few if any minimal pairs in normal speech. A systematic use of pitch in Cantonese English was reported in recent studies such as [5, 7-10]. This paper refers to this systematic use of pitch as tone. It asks whether each 'tone' is significantly different from each other phonetically, and how some tones span across several syllables forming tone spans. This paper examines these newly developed tones by fitting a smoothing spline analysis of variance (SS ANOVA) generated from R (version 3.0.2) [11] to the acoustic data and showing the pitch tracks of different tone spans from Praat (version 5.3.39) [12]. Based on the phonetic analysis, it provides generalizations of the phonetic distribution of tones in Cantonese English and illustration of tonal associations.

2. Phonetic Evidence

Audio-recordings were obtained from six speakers of Cantonese English. They were balanced for biological gender, all born after 1980, and raised in Hong Kong. The data was elicited with a wordlist covering the surface tones (M(id), H(igh), M(falling), L(ow), H(f) and corner vowels ([i], [a], [u]) of Cantonese English. The speakers were asked to produce each word three times in a row so that the basic position (initial, medial or final) of each target item in an utterance was controlled, for instance, elicit-eliciting. The wordlist was repeated twice for taking averages. A total of 2376 target syllables (= 66 target syllables x 3 repetitions x 2 sets x 6 subjects) were recorded using Praat with a sampling frequency of 22050Hz in a sound-proofed recording booth.

Fundamental frequencies (F0) of the rhymes of each word in the utterance final position were extracted with Praat, then time-normalized at 10% interval points with ProsodyPro (version 4.3) [13]. The F0 was then transformed into logarithmic Z-score (LZ-score) values to fit the logarithmic characteristic of pitch in speech perception [14] and production [15], and to minimise inter-speaker variation. The conversion was based on equation (1):

$$LZ-score = \frac{\log x - \log \bar{x}}{\log SD}$$

(1)

where x is F0, \bar{x} is mean of F0, and SD is standard deviation.

Now we are ready to fit a SS ANOVA to the LZ-scores to see if the tones in different word positions differ significantly from each other.

2.1. SS ANOVA plots

A SS ANOVA was implemented to the LZ-score values with R. It is a statistical technique used to estimate the means of dependent variables having a contour characteristic over a certain dimension, in this case time, with an ANOVA [16]. A smoothing spline of the estimated values is constructed by connecting discrete data points with a polynomial function which concerns the mean, error effect, shape of the curve, and interaction between the effect and shape of the curve. This technique has been used to determine whether the shapes of multiple curves statistically differ from one another significantly in articulatory phonetics when comparing tongue shapes on ultrasound images [17-19] and in acoustic phonetics for tonal contours in F0 [20-22]. This study applied a SS ANOVA to the F0 data to generate the contours which show the smoothing spline fit for each of the tones.

To confirm whether the smoothing splines of the LZ-scores are statistically different, 95% confidence intervals (CIs) were constructed for the sets of curves. If the CIs at any section on the length of two curves overlap, the difference between them in that section is not statistically significant. A wider CI at a certain interval point of time indicates a greater variance in the LZ-scores of all data available at that point.

Figure 1 shows the tone profile of Cantonese English for speaker F2 as an example. Recall that the words were repeated three times so as to control the position of the words in an utterance. Only the words in the utterance-final position are presented because they display richer materials for tonal analysis. Also, since the placement of tones seems to be restricted to syllable positions, three graphs, from left to right, are used to present the tones found on syllables in word-initial, word-medial and word-final positions respectively. A word-initial syllable refers to the first syllable of a word while the word-final syllable refers to the last syllable of a word. Any remaining syllables in between are treated as word-medial. The curves with the highest LZ-scores are H whereas those with the lowest ones are L, except for the initial position.
where the curve with the lower LZ-score is M. The one between H and L in the medial position is also M.

As shown, not all tones surface in all syllable positions. H and M surface in both word-initial and word-medial positions. In the word-final position, the lower curve is regarded as a boundary L, and the upper curve with a distinguishable falling contour is a H transiting to the boundary L. Additional data in section 2.3 shows that the more the syllables between the rightmost H and the boundary L, the flatter the slope of each curve. They are interpolating tones from the rightmost H in a word to the boundary L.

Since the curves at 95% CIs do not overlap with each other, their difference is statistically significant. We are confident that the use of pitch as tones is systematic and stable in the way mentioned above. Despite some variations concerning the details of the pitch contours, the patterns described above generally hold across subjects.

The sequence of tones is identified as follows:

\[(M \ (\text{span})) - H \ (\text{span}) - (\emptyset \ (\text{span})) - %L\]

There are three kinds of tone spans: H spans, Ø spans and M spans. Each of them will be illustrated with pitch tracks of individual words as examples in sections 2.2, 2.3 and 2.4.

### 2.2. H spans

Two kinds of H spans were recognized from the data. The Hs of one type surface as H in words of all three utterance positions while the Hs of the other type only surface in non-utterance-final positions. H spans in ‘understood’ and ‘criminal’ are examples of the two types respectively.

Figure 2 displays two plots of pitch tracks. F0 is based on Hertz scale with time in seconds. From top to bottom, it shows the pitch track tier, syllable tier and tone tier. The dashed lines represent the syllable boundaries of the words. Pitch tracks were extracted from the same speaker for convenience of comparison of tones.

Figure 3 introduces a phonetic H span. Syllables following the H on the first syllable of a word are found to have alternating tonal behaviour depending on the position of the word in an utterance. The last two syllables of ‘criminal’ surface as H in non-utterance-final position, but as Mf and L in utterance-final position. This alternating tonal behaviour suggests that the last two Hs are phonetic Hs, forming a phonetic H span. The phonetic Hs receive their tone value from the only tone in the same word, i.e. the H on the first syllable. When the word is at the utterance-final position, the last syllable will receive a boundary L, and the syllable between the phonological H and the boundary L transits from the H to reach the L target. Considering the alternating behaviour of the tones on the syllables following the phonological H, let us call those two alternating syllables Ø syllables, and move on to section 2.3 for more details.
2.3. Ø spans

Focusing on the tonal contour of the Ø syllables, Figure 4 shows an example with three Ø syllables.

Figure 4: Ø syllables of 'regularly' in non-utterance-final position (upper panel) and utterance-final position (lower panel).

Similar to 'criminal', for 'regularly', other than the first syllable which carries H no matter the word is in the utterance-final position or not, all three remaining syllables alternate as either H when the word is not at the utterance-final position, or Mf/L when the word is at the utterance-final position. One more Ø syllable in 'regularly' than in 'criminal' for tonal interpolation leads to a flatter slope of the Mf spans in 'regularly'. This indicates that the more the number of syllables available between the phonological H and the boundary L, the flatter the slope of those syllables. Ø syllables seem to be phonologically toneless and can carry any interpolating tones demanded by the utterance boundary.

2.4. M spans

M spans are found preceding the phonological H (span) in a word. Figure 5 shows an example.

For 'introduce', there is a M span preceding the Hf. Syllables preceding Hf surface as M in all three utterance positions, unaffected by any positional effects. When comparing the pre-high M (as the first two Ms in 'introduce') and post-high Mf (as the last two Mfs in 'criminal'), it is clear that they are different in terms of their response to positional/boundary effects and their shapes. The pre-high M is phonetically very stable, unaffected by any positional/boundary effects while the post-H Mfs are highly responsive to boundary effects. When it comes to their shapes, the pre-high Ms are more horizontal (cf. Figure 5) whereas the post-high Mfs are either steeper (cf. Figure 3) or flatter (cf. Figure 4) depending on the number of post-high phonologically toneless syllables available for tonal interpolation to reach the target boundary L. Therefore, though the post-high Mfs are interpolating tones from H to L, they are not necessarily M.

3. Generalizations and association of tones

Based on the discussion of phonetic data, here are some generalizations of the tones in Cantonese English.

- H is obligatory. There is at least one H in each lexical word. Its high feature spreads rightward to all post-high Ø syllables in a non-utterance-final word.
- M is optional, only found preceding the obligatory H (span) of a word. It is phonetically very stable in terms of shape, unaffected by any boundary effects.
- Ø is optional, only found following the obligatory H (span) of a word if any. It surfaces as either H or Mf/L depending on whether the word is utterance final.
- There are three types of tone spans: H spans, Ø spans and M spans.
- H spans can be phonological or derived from spreading by the obligatory H of a word.
- Ø spans are a chain of interpolating H or Mf/L. In an utterance-final word, the Ø span surfaces as phonetic Hs. In an utterance-final word, the slope of the pitch contours of the Mf spans correlates inversely with the
number of syllables available between the phonological H and %L of the word.

- M spans are phonetic defaults grounded on physiological case. The stabilities of the pitch contour shape and neutrality to positional/boundary effects suggest that M may also be carrying some sort of phonological function.

Tonal association of different spans will be illustrated assuming syllable as the TBU. Solid lines and dashed lines indicate phonological tones and derived tones respectively.

### 3.1. H spans

![Figure 6: Phonological H spans (upper panel) and phonetic H span (lower panel).](image)

The H span of ‘understood’ results from associating the obligatory H with all three TBUs, having three syllables sharing a H. Since there is a phonological H on the utterance-final syllable, the %L has to share a TBU with that H, resulting in a Hf. The H span of ‘criminal’ is derived from spreading of the obligatory H to the phonologically toneless resulting in a Hf. The H span of final syllable, the %L has to share a TBU with that H, obligatory H with all three TBUs, having three syllables depending on the utterance boundary. There is a chain of spreading of the obligatory H to the L target without associating with any tones.

### 3.2. Ø spans

![Figure 7: Ø spans.](image)

Figure 7 presents another example with one more intervening syllable between the obligatory H and the %L. Similarly, the Ø syllables alternate between phonetic Hs and M/L depending on the utterance boundary. There is a chain of phonetic Hs, or interpolating tones from the obligatory H to the %L.

### 3.3. M spans

![Figure 8: M spans in all three utterance positions.](image)

The M spans of ‘introduce’ result from associating the M with two adjacent TBUs. The fact that the Ms in the utterance-medial word do not receive Hs from the preceding word could be due to the requirement of alignment between the left edges of the syntactic boundary and the phonological boundary [9], which is marked by the pre-high M(s) on the left edge of a phonological phrase, i.e. a word in this case. It serves as a phonological boundary, preventing spreading from the preceding H in another phonological phrase to the present M.

From the above, we can see that a tone can be linked to multiple TBUs and a TBU can be linked to multiple tones in Cantonese English.

### 4. Conclusions

This paper has examined the newly developed tones in Cantonese English with a particular interest in the tones spanning across syllables based on the acoustic data. It displayed a distinctive use of pitch Cantonese English, and illustrated the H spans, Ø spans and M spans with pitch tracks and association of tones to TBUs.

### 5. Acknowledgements

I thank Stephen Matthews and Diana Archangeli for many stimulating discussions. I also thank Lianhee Wee, ‘lab rats’ Winnie Cheung and Queenie Chan, and the six subjects. This paper also benefits from useful comments by two anonymous reviewers and the audience at TAL4.

### 6. References


