Intonation of statements and questions in Cantonese English: acoustic evidence from a smoothing spline analysis of variance

Suki Yiu

The University of Hong Kong
syutji@hku.hk

ABSTRACT

This paper looks into the intonational use of pitch in Cantonese English. It aims at providing phonetic evidence for the intonation of statements and questions in interaction with tones identified at lexical level. A smoothing spline analysis of variance was fitted to the audio-recording data from six speakers of Cantonese English with R in order to generate smoothing splines at 95% confidence intervals for determining whether the same tone (M, H, or L) in different utterance types (context-neutral utterances, statements and questions) differ from each other significantly in different word positions and utterance positions. It is found that there are clear boundary targets for statements (L%) and questions (H%), indicating that there are different boundary tones. It also finds out that there is a lowering of pitch for tones in non-utterance-final words for questions than other utterance types, which creates an enhancing effect of a more salient H%.

Keywords: intonation, tone, Cantonese English, Hong Kong English, SS ANOVA

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 [15]. In fact, many aspects of the phonology of Cantonese English have been studied so far [10-12, 17, 19]. Specifically, a systematic use of pitch in Cantonese English has been reported in recent studies such as [3, 9, 12, 20, 22-23]. This paper refers this systematic use of pitch as tone. It asks whether different utterance types can be reflected by certain ways of pitch manipulation.

In order to answer the above question, I processed the acoustic data with Praat (version 5.3.39) [1], and then fitted a smoothing spline analysis of variance (SS ANOVA) with R (version 3.0.2) [18] to the data. The smoothing splines representing pitch contours of different tones in context-neutral and contextual utterances were compared with each other for intonation in terms of fundamental frequency (F0). Context-neutral utterances refer to utterances produced without a syntactic context, whereas contextual utterances refer to those with a syntactic context such as statements and questions. It is hypothesized that there is an intonational use of pitch in Cantonese English. This predicts that the smoothing splines in an utterance are modified in a systematic way depending on utterance types. If the smoothing splines at 95% confidence intervals (CIs) of the contextual utterances cannot be fully covered by those of the context-neutral ones, there is intonational use of pitch displaying phonological characteristics of tone assignment to be analysed.

2. METHODOLOGY

Audio-recordings were collected 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), Mf(alling), L(ow), Hf) of Cantonese English.

The tones in a context-neutral utterance were elicited by asking the subjects to produce the target words three times in a row so as to obtain F0 information for three utterance positions, i.e. initial, medial and final, for example, *criminal-criminal-criminal*. Also, choosing words with different tonal distribution over a word would provide F0 information of certain tones in different word positions. Let us take *criminal-criminal-criminal*, *elicit-elicit-elicit* and *disagree-disagree-disagree* as examples. Focusing on the utterance-initial word, the three underlined syllables in *cri*minal-criminal-criminal, *e*licit-elicit-elicit and *disa*gree-disagree-disagree give F0 data of H in the three word positions respectively. By also getting the F0 of H in the three word positions of utterance-medial and utterance-final words, the F0 trajectory of one of the target tones, H, in the whole utterance could be obtained.

The tones in contextual utterances were elicited through statements and questions. Target words were placed in syntactic frames of the two utterance types with three utterance positions and three word positions. The same target words for context-neutral utterances were adopted here in order to make direct

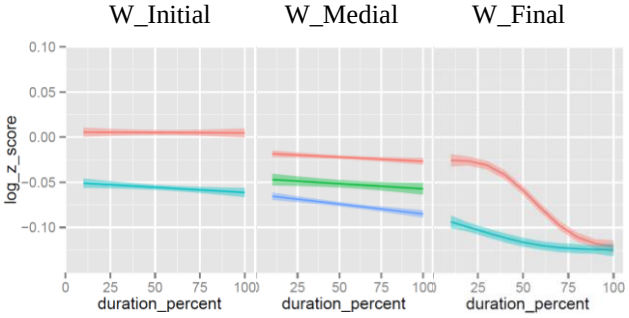
comparison between the smoothing splines of the context-neutral and contextual utterances. The list of words was recorded twice to calculate mean values at different time intervals. During the data elicitation process, prompt questions were asked to mimic a natural exchange in normal speech.

A total of 7128 target syllables (= 66 target syllables x 3 utterance positions x 3 utterance types x 2 sets x 6 subjects) were recorded using Praat (ver. 5.3.39) [1] at a sampling frequency of 22050Hz in a sound-proofed recording booth.

F0 of the rhymes of each word were extracted with Praat, then time-equalized at 10% interval points with ProsodyPro (version 4.3) [21]. The F0 was then transformed into logarithmic z-score (LZ-score) values to fit the logarithmic characteristic of pitch in speech perception [16] and production [7], and to make inter-speaker comparison. After that, a SS ANOVA was fitted to the LZ-scores to see whether the smoothing splines representing different tones differ significantly from each other in different word and utterance positions.

The SS ANOVA implemented to the LZ-score values is a statistical technique used to estimate the means of dependent variables having a contour characteristic over a certain dimension with an ANOVA [8]. A smoothing spline of the estimated values is constructed by connecting discrete data points with a polynomial function. This technique has been used to determine whether the shapes of multiple curves are significantly different from one another statistically in articulatory phonetics when comparing tongue shapes on ultrasound images [2, 5, 6], and more recently in acoustic phonetics for tonal contours in F0 [4, 13-14, 23]. 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 in different word and utterance positions. For example, Figure 1 shows the SS ANOVA plots of the utterance-final word for speaker F2.

Figure 1: SS ANOVA plots of the utterance-final word for speaker F2.



The contours are smoothing spline fits for each tone on syllables in different word positions. 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 shaded region along each spline is the 95% CIs. The splines with the highest LZ-scores are H whereas those with the lowest ones are L, except for the initial position where the spline 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 word positions. H and M surface in both word-initial and word-medial positions. In the word-final position, the lower spline is regarded as a boundary L (L%), and the upper spline with a distinguishable falling contour is a H transiting to the L%. [23] showed that the more the syllables between the rightmost H and the L%, the flatter the slope of each spline. They are interpolating tones from the obligatory H in a word to the L%, and are referred to as Ø henceforth.

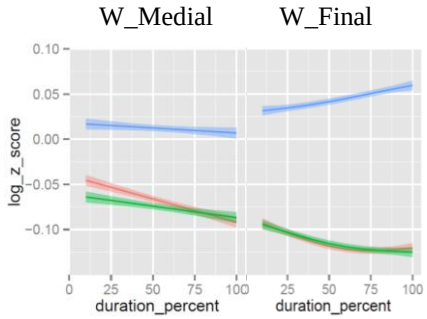
3. TONES IN CONTEXT-NEUTRAL AND CONTEXTUAL UTTERANCES

This section compares the SS ANOVA plots of the same tone in two types of utterance: context-neutral and contextual (statement and question) in order to see whether the utterance type has any effect on the use of pitch at intonational level in Cantonese English. These plots visualise the splines of each tone, with the utterance types as the variable. If the smoothing splines of the statements and questions cannot be entirely covered by those of the context-neutral utterance at 95% CIs, the non-overlapping regions in statements and questions are interpreted as an extra use of pitch demanded by the contextual information to be conveyed, that is, intonation.

3.1. Utterance boundary tones on Ø syllables in utterance-final words

From Figure 2 onwards, the splines of different colours represent tones of different utterance types: the green, red and blue splines represent tones of context-neutral utterances, statements and questions respectively. The colour splines in Figure 2 display the pitch contours of Ø syllables in utterance-final words.

Figure 2: Ø syllables in utterance-final words for speaker F2.



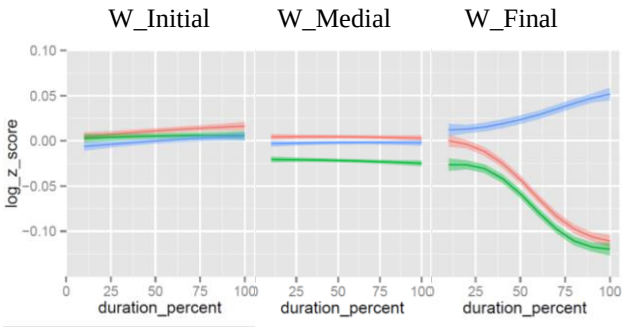
In both word-medial and word-final positions, the blue splines of questions are significantly way above the other splines, indicating that there is intonation for questions. There is a H target at the utterance-final boundary for the Ø syllables in the utterance-final words in questions to reach. This pattern holds for all subjects.

In the meanwhile, the red splines for statements barely differ significantly from the green splines for context-neutral utterances. For the word-final Ø syllable, the red splines of statements for all six subjects are covered by the green splines of context-neutral utterances for at least 50 percent of the splines in different ways: The second half of the red spline of statements for speaker F2 and the first half of that for F3 are covered by their respective green spline of context-neutral utterances; Speakers M1 and F1 have their respective upper half and lower half of the red statement spline overlapping with the green context-neutral spline; Speaker M2 displays a total overlap. For the word-medial Ø syllable, speaker F2 has the red splines of statements and green splines of context-neutral utterances almost entirely overlapping with each other. Speakers M1, M2, M3 and F1 have their red splines of statements just above their green splines of context-neutral utterances. Speaker F3 shows the opposite. One way to explain why the splines of statements tend to overlap with the splines of context-neutral utterances is that context-neutral utterances are not entirely intonationally neutral, but still carrying an intonation of listing, which is closer to the intonation of statements. We will see more phonetic evidence of a similar overlapping situation for the Ø syllable in utterance-initial and utterance-medial words in section 3.3.

3.2. Utterance boundary tones on H in utterance-final words

Figure 3 shows the manifestation of utterance boundary tones in another type of words where no Ø syllable is available in the word-medial and word-final positions for F0 to interpolate to the target boundary tones in utterance-final words.

Figure 3: H in utterance-final words for speaker F2.



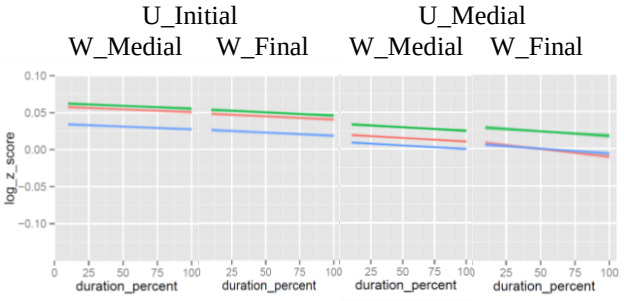
The word-final syllable in Figure 3 displays F0 transitions from a H onset to the respective red L% offset or blue H% offset. The boundary tones share the word-final syllable in the utterance-final word with the phonological H. For speaker F2, we can see a clear split in the utterance-final syllable between the red statement spline and blue question spline at around 15 percent on the time dimension. The F0 of the word-final syllable in questions climbs towards an even higher target, i.e. H%, creating a rising contour, whereas the F0 in statements falls to meet the L%, creating a falling contour. Other speakers share the same pattern but vary in terms of the exact point of divergence in the utterance-final syllable.

Figure 3 has displayed another way in which the boundary tones of statements and questions can surface on words having a phonological H on the utterance-final syllable. The use of boundary tones in words with and without Ø syllables display a use of intonational pitch in Cantonese English.

3.3. F0 overshooting of Ø syllables

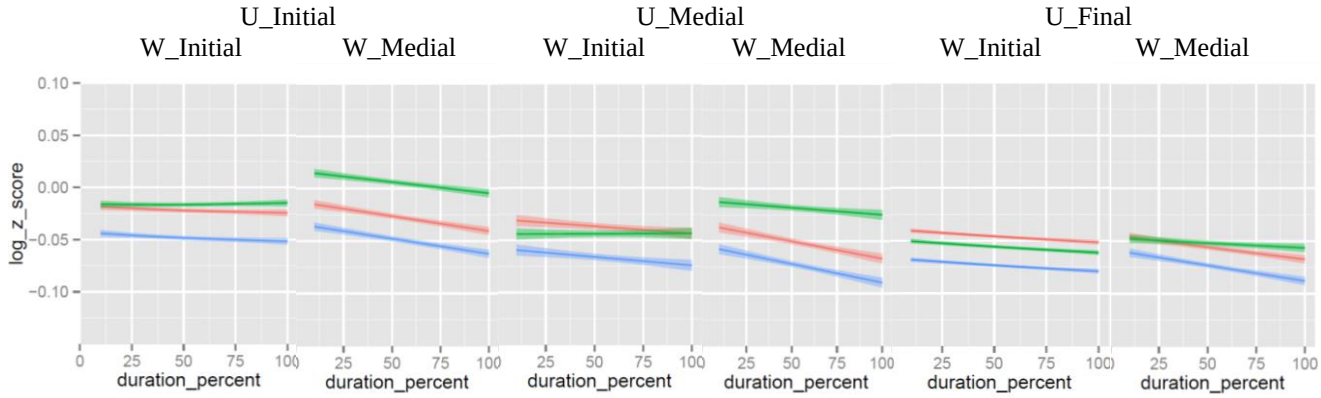
In addition to the boundary tones, another acoustic cue signalling utterance types is the F0 overshooting of Ø syllables in non-utterance-final words.

Figure 4: Ø syllables in non-utterance-final words for speaker F2.



In Figure 4, the blue splines of utterance-initial and utterance-medial words for questions are

Figure 5: M in all word and utterance positions for speaker F2.



consistently lower than the red splines of statements and green splines of contextual utterances, contrary to those for utterance-final words in Figure 2. This pattern is found for all six speakers, except for speaker F1, where the blue spline of the word-medial syllable in the utterance-initial word is found above the red and green splines around the 40 percent time-interval point. The reason why the splines of Ø syllables for questions start lower in the utterance-initial and utterance-medial words could be to highlight the contrast of the H% at the end of the utterance. Since the Ø in non-utterance-final words for all utterance types surface as phonetic H, lowering this phonetic H can enhance the H% or rising effect on the whole. This makes sense for the articulatory control on the part of the speakers to achieve a desirable perceptual effect for the listeners. We will see more of this enhancing effect in Figure 5 for M.

In particular, words with and without Ø syllable display some small difference in where the enhancing effect ends. The enhancing effect of words with Ø syllables from Figures 2 and 4 last till the utterance-medial words for all speakers. While for words without Ø syllables in Figure 3, since the divergence does not happen until the utterance-final word, the blue question splines are allowed to stay lower than the red statement splines up to the word-medial syllable, or even part of the word-final syllable, of the utterance-final words. This observation adds details to how the enhancing effect works for boundary tones.

Besides, similar to the overlapping situation of utterance-final words in Figure 2, the green splines of Ø for context-neutral utterances in utterance-initial and utterance-medial words in Figure 4 overlap more with those red splines of statements than with those blue splines of questions. As mentioned in section 3.1, this may due to the listing intonation of the context-neutral utterances.

3.4. M in non-word-final syllables

To complete the picture of how intonation is realised in interaction with the tones at lexical level, let us look at syllables which surfaces as M in Figure 5. It shows that the blue splines of questions are consistently lower than the red splines of statements for speaker F2, owing to the enhancing effect for the boundary tones which is also found for Ø syllables (cf. Figures 2 and 4) and H (cf. Figure 3). The same pattern is found for other speakers. Moreover, the blue splines of questions do not overlap with those green splines of context-neutral utterances, suggesting that the difference between them is statistically significant. Though part of the red splines of statements are covered by the green splines of context-neutral utterances, that the red splines of statements cannot be fully covered in the whole utterance still supports that the M splines of statements and context-neutral utterances are different.

4. CONCLUSIONS

In this paper, I have demonstrated the use of a statistical tool newly adopted in acoustic phonetics, SS ANOVA, to illustrate different boundary targets for statements and questions. A close examination of the splines over the whole utterance reveals the lowering of pitch for tones in non-utterance-final words for questions to create an enhancing effect of a more salient H%. Since the splines for questions do not overlap with those for context-neutral utterances, there is use of intonation for questions. Some splines for statements are covered by those for context-neutral utterances. As explained, the overlapping may be due to the fact that context-neutral utterances still carry an intonation of listing, which is similar to the intonation of statements. Given the above findings, I conclude that there is an intonational use of pitch for both statements and questions in Cantonese English.

5. ACKNOWLEDGEMENTS

I sincerely thank Stephen Matthews and Diana Archangeli for various motivating discussions. Many thanks also go to Jon Yip, Lianhee Wee, Cathryn Donohue, and the six subjects.

6. REFERENCES

[1] Boersma, P., Weenink D. 2013. Praat: doing phonetics by computer [Computer program]. Version 5.3.57, retrieved 2012 from http://www.praat.org/.

[2] Chen, Y., Lin, H. 2011. Analysing tongue shape and movement in vowel production using SS ANOVA in ultrasound imaging. *Proc. of 17th ICPhS*, 123-127.

[3] Cheung, W. 2009. Span of high tones in Hong Kong English. *Proc. of BLS 35*, 72-82. Berkeley, CA, 2009.

[4] Chuang, C.T., Chang, Y.C., Hsieh, F.F. 2013. Complete and not-so-complete tonal neutralization in Penang Hokkien, In: Lee, W.S. (ed.), *Proc. of ICPLC 2013*, 54-57.

[5] Davidson, L. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America* 120 (1): 407-415.

[6] Fruehwald, J. SS ANOVA. 2010. http://www.ling.upenn.edu/~joseff/.

[7] Fujisaki, H. 2003. Prosody, information, and modeling: With Emphasis on Tonal Features of Speech. *Proc. of Workshop on Spoken Language Processing*, Mumbai, 5–14.

[8] Gu, C. 2002; 2013. *Smoothing Spline ANOVA Models*. Springer, New York.

[9] Gussenhoven, C. 2012. Tone and Intonation in Cantonese English. *TAL 3*, paper O3-04.

[10] Hung, T. 2000. Towards a phonology of Hong Kong English. *World Englishes* 19(3), 337-356.

[11] Hung, T. 2005. Word Stress in Hong Kong English: a preliminary study. *HKBU Papers in Applied Language Studies* 9, 29-40.

[12] Luke, K.K. 2000. Phonological re-interpretation: The assignment of Cantonese tones to English words. Paper presented at *ICCL 9* NUS.

[13] Mathes, T. 2014. Extreme tonal depressor effects in Khoisan: Evidence from Tsua. In: Gussenhoven C., Chen, Y., Dediu, D. (eds.), *Proc. of TAL 4*, 1-5.

[14] Moisik, S., Lin, H., Esling, J. 2013. Larynx height and constriction in Mandarin tones. In: Peng, G., Shi F. (eds.), *Eastward flows the Great River: Festschrift in Honor of Professor William S-Y. Wang on his 80th birthday*, 187-205.

[15] Mohanan, K.P. 1992. Describing the Phonology of Non-Native Varieties of a Language. *World Englishes* 11, 111-128.

[16] Nolan, F. 2003. Intonational equivalence: an experimental evaluation of pitch scales. *Proc. of 15th ICPhS*, 771-774.

[17] Peng, L., Setter, J. 2000. The Emergence of Systematicity in English Pronunciations of Two Cantonese-speaking Adults in Hong Kong. *EWW* 21(1), 81-108.

[18] 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/.

[19] Sewell, A., Chan J. 2010. Patterns of variation in the consonantal phonology of Hong Kong English. *EWW* 31(2), 138-161.

[20] Wee, L.H., 2008. Phonological patterns in the Englishes of Singapore and Hong Kong. *World Englishes* 27(3/4):480-501.

[21] Xu, Y. 2012. Praat script ProsodyPro (version 4.3).

[22] Yiu, S. 2010. Intonation of English Spoken in Hong Kong. BA Thesis, Hong Kong Baptist University.

[23] Yiu, S. 2014. Tone spans of Cantonese English. In: Gussenhoven C., Chen, Y., Dediu, D. (eds.), *Proc. of TAL 4*, 143-146.