CONTOUR SHAPE AND PROSODIC FOCUS MARKING: THE CASE OF HONG KONG CANTONESE

Chris K. C. Lee & Jonathan Barnes

Linguistics Department, Boston University
chriskcl@bu.edu, jabarnes@bu.edu

ABSTRACT

Hong Kong Cantonese has been reported to show inconsistent, if not absent, pitch-related modulation in response to changes in information structure. However, previous studies focused primarily on F0 target scaling, neglecting other important factors, such as the overall shape of the F0 contour. This study focuses on the High level (T55) and High rising tones (T25), realized between a preceding Low level T22 and a following Low falling T21, and investigates whether Cantonese speakers modulate the contour shapes of the High tones in implementing information-structural contrasts. Exploiting *Tonal Center of Gravity in the time dimension* (TCoG), we confirm that Cantonese speakers hyperarticulate each tone's characteristic shape under focus. When flanked by two Low targets, on-focus T55 have a significantly domier rise, and on-focus T25 a significantly scoopier rise, than their respective broad-focus counterparts, enhancing the contrast between the High tones in terms of TCoG.

Keywords: Prosodic focus marking; Tonal Center of Gravity; Hyperarticulation; Cantonese

1. INTRODUCTION

Information structure is expressed in various ways in the world's languages [1], [2]. Phonetically, constituents with narrow focus are often realized with a wider pitch range, longer duration, and/or higher intensity than those in broad focus, and post-focus constituents with compressed pitch range and intensity (e.g. English [3]–[5], German [6], Mandarin [7], Japanese [8]). Nevertheless, how and whether focus is marked prosodically remains language-specific in nature. In Cantonese[1], longer constituent duration has been argued to be the only consistent cue to narrow focus [9]–[11]. Pitch range of focused constituents is reported to be modulated only inconsistently: while [12], [13] present evidence for expanded pitch range in contrastively focused constituents, [11] reported that in Cantonese on-focus expansion of pitch range is only observed in the two rising tones, while pitch level raising is observed inconsistently in the non-High level tones. [11], [12] also report that post-focus compression (PFC), whether of pitch, intensity, or duration, is non-existent in Cantonese.

The studies on Cantonese cited above, however, all rely on F0 measures such as mean, range, or extremum level within a region of interest. While these measures characterize at least in part the *scaling* of a pitch event, using *only* these measures neglects two other important prosodic properties known to interact with target scaling perceptually: *timing* and *contour shape* of a pitch event [14]–[22]. Concerning the timing of pitch events, [7, p. 86] reports that in Mandarin a rising tone followed by a Low target has *later* F0 peaks when under focus than those not under focus. Timing differences of this kind, however, are also often accompanied by a contour shape difference [16]. In the context of a high level tone, for example, aligning an F0 peak earlier would also mean staying at the max F0 for longer within the tone-bearing syllable, creating a longer High plateau.

Since focus-induced modulation of F0 contour shape of Cantonese tones is hitherto unexplored, this paper aims to investigate

RQ1: whether Cantonese speakers vary the contour shape of High lexical tones in realizing different information structures, and

RQ2: whether tone type affects the direction and magnitude of contour shape modulations induced by focus.

Instead of tracking the timing or scaling of any single "target" point in the F0 contour, this study utilizes the *Tonal Center of Gravity in the time dimension* (**TCoG**), a measure that estimates when a pitch event happens *perceptually*, to characterize the perceptual consequences of a change in contour shape [15]–[17]. TCoG uses weighted F0 averages to generalize about the overall disposition of a pitch event in time. For a High tonal target associated with a given syllable, earlier F0 peaks, higher pre-High F0 valleys, longer High plateaux, and/or domier rises will shift the bulk of the High F0 region, and thus TCoG, leftward. Later F0 peaks, lower pre-High F0 valleys, and/or scoopier rises, by contrast, shift TCoG rightward. Fig. 1 schematizes these generalizations. (See [16] for discussion.)

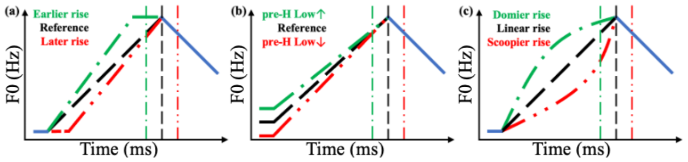

Figure 1: Schematics for the effect of (a) pitch event timing, (b) pre-High F0 valley scaling, and (c) contour shape curvature on TCoG.

Since focus-induced variation in High target timing was reported when it precedes a Low target in a tone language like Mandarin [7], and the applicability of TCoG to characterize the contour shape of a L-H-L tonal sequence is well attested, this paper focuses on Cantonese High tones (T55, T25) surrounded by Low targets (T22, T21).

2. METHOD

In this study, participants read a set of sentences with a natural and context-appropriate tone of voice as a response to some contextualizing questions pre-recorded by a male native Cantonese speaker in his late 20s born and raised in Hong Kong.

2.1. Stimuli

The elicitation materials of the present study are created in such a way that a High pitch target is surrounded by Low targets. The target NPs this study adopts are four trisyllabic personal names with one of the following tone sequences (Fig. 2 shows the pitch tracks of the two tone sequences):

T22-55-21 (*T55 set*): 鄭依明 [tsɛŋ^{22} ji^{55} mɪŋ^{21}],
段威龍 [tyn^{22} wɐj^{55} lʊŋ^{21}]
T22-25-21 (*T25 set*): 范綺玲 [fan^{22} ji^{25} lɪŋ^{21}],
鄧委文 [tɐŋ^{22} wɐj^{25} mɐn^{21}]

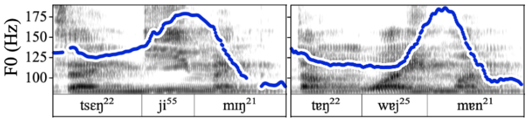

Figure 2: Pitch tracks of a T22-55-21 (left) and a T22-25-21 (right) target NP produced in broad focus condition by a male participant in the current study.

The target NPs are embedded in the carrier sentence NP$_1$約NP$_2$過大海 [*NP$_1$* jœk̚^{3} *NP$_2$* k^{w}ɔ^{33} taj^{22} hɔj^{25}] (*NP$_1$ invited NP$_2$ to visit Macau*). Each target NP appears in both NP$_1$ and NP$_2$ positions, but in each sentence NP$_1$ ≠ NP$_2$. This generates 4×(4 – 1) = 12 variations of the carrier sentence.

As for the contextualizing questions, they are designed to elicit different contexts such that NP$_2$ is placed in broad-focus (1), post-focus (2), narrow-focus (3), pre-focus (4) or contrastive-focus conditions (5).

(1) What's new recently?
(2) Who invited NP$_2$ to visit Macau?
(3) Whom did NP$_1$ invite to visit Macau?
(4) What did NP$_1$ invite NP$_2$ to do?
(5) Did NP$_1$ invite [siw^{22} jɪŋ^{55} mɪŋ^{21}] to visit Macau?

2.2. Acoustic analysis

For each sentence, only NP$_2$ was analyzed. Each participant produced 100 trials (4 target NPs × 5 focus conditions × 5 repetitions). Boundaries of the target NPs and syllables therein were annotated in Praat [23]. Acoustic measures, including F0, NP duration (onset of the first syllable excluded), and (root-mean-squared) energy were estimated with VoiceSauce [24] with REAPER [25] integrated for F0 estimation. Default settings were adopted. All measures are *z*-transformed within speaker.

TCoG was used to identify the perceptual reference location in the time dimension for the High target associated with the target NP-medial syllable, in order to investigate the influence of the rising contour shape on the perception of tonal target timing. Its calculation was based on the basic formula proposed in [16], reproduced in (6), where F0$_i$ is the F0 at time t$_i$ relative to the min. F0 of the region of interest.

$$\text{(6)}\quad \text{TCoG} = \frac{\Sigma_i F0_i \times t_i}{\Sigma_i F0_i}$$

The start and end points of the region of interest corresponded to the F0 valley before the F0 peak of a target NP and the F0 peak respectively. To compare TCoG values across focus conditions, we took the midpoint of the target NP-medial syllable as a segmental anchor.

3. RESULTS

Data from 13 native Cantonese speakers (six male and seven female university degree holders, aged 24–30 [*M*=27.92], born, raised, and educated in Hong Kong) were analyzed. Among them one speaker merged T25 with the low rising tone completely, so all of their T25 tokens (*N*=50) were discarded. 22 more T25 tokens and three T55 tokens of the other speakers were discarded due to segmental or tonal errors or disfluency. Altogether, 647 tokens in the T55 set and 578 tokens in the T25 set were analyzed.

3.1. The "conventional" acoustic measures

We first consider the more well-studied acoustic correlates of focus marking in Cantonese (NP duration, mean energy) to ensure our participants express information structure in the expected direction. To assess the effect of focus on pitch scaling, we also consider the target NPs' mean F0 and max F0 of the window for TCoG calculation. Effectively, these two measures quantify the pitch scaling of the High pitch target. Fig. 3 shows the group means and standard errors of these values separated by focus condition. Values of broad-focus NPs are represented by red vertical lines. To statistically assess our observations reported below, separate linear mixed effects models (LMEMs) were fitted, using the four measures as dependent variables. We considered FOCUS, TONESEQ (T55 set *vs.* T25 set), and their interaction as potential fixed effects, and included a random intercept of speaker and a random intercept of target NP.[2] Based on all-subset selection, the best models for all four acoustic measures select FOCUS as the only significant main effect.

Let's start with on-focus marking. Compared to broad-focus NPs, narrow- and contrastive-focus NPs (light and dark green respectively) are realized with significantly longer duration (Narrow-focus [NF]: β=1.20, t=18.79, p<.001; Contrastive-focus [CF]: β=1.21, t=19.01, p<.001) and higher mean energy (CF: β=.225, t=5.48, p<.001; NF: β=.230, t=5.58, p<.001). For pitch scaling measures, our participants realize NPs in different focus conditions with statistically comparable max F0 (ps>.60). As for mean F0, although Fig. 3 suggests that mean F0 is lowered slightly when the NP is under focus, such a difference is statistically not significant (ps>.055). Our participants are hence very comparable to those in previous studies in terms of on-focus marking.

The more interesting finding here concerns PFC where, contrary to previous findings, participants actively compress NP duration (β=-.427, t=-6.73, p<.001), mean energy (β=-.196, t=-4.78, p<.001), and the pitch scaling of the High target (max F0: β=-.524, t=-12.14, p<.001; mean F0: β=-.325, t=-10.22, p<.001) of post-focus NPs relative to broad-focus NPs. In fact, while PFC in general is smaller in magnitude than on-focus expansion, PFC is consistently observed in all the acoustic measures that previous studies have considered.

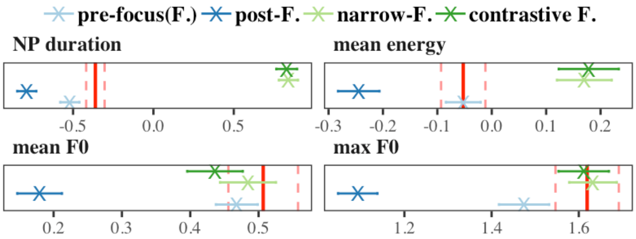

Figure 3: Means of the four "conventional" acoustic correlates for prosodic focus marking in Cantonese. Red solid lines indicate parameter values of broad-focus NP. Error bars and dashed lines indicate SE.

3.2. Contour shape

Apart from the somewhat surprising appearance of PFC in our data, participants otherwise produced focus-related modulations of acoustic properties in expected directions. We now consider the contour shape. Fig. 4 shows time-normalized F0 tracks from the second half of the NP-initial syllable of the target sequence to the first 40% of the NP-final syllable averaged across speakers, separated by tone set. Once again corroborating our findings with respect to PFC of pitch scaling, post-focus NPs (dark-blue dashed line) have a much lower pitch level than NPs in other focus conditions. However, its contour shape is quite comparable to broad-focus NPs (red solid line).

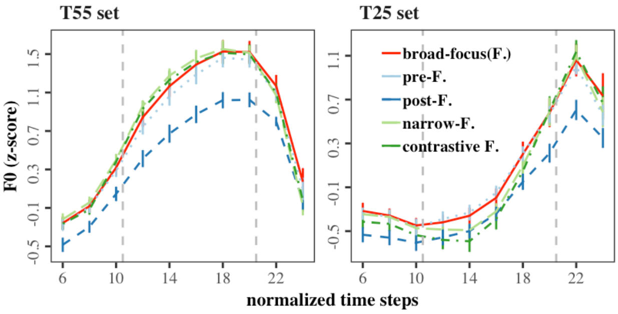

Figure 4: Time-normalized F0 contour of NPs by focus condition and tone set averaged across speakers. Each syllable in NPs is represented by 10 equidistant points.

On the other hand, the contour shapes of narrow- and contrastive-focus NPs (light green dashed and dark green dot-dashed lines respectively) are quite different from the broad-focus variants, and the direction of focus-induced contour shape modulation is tone-sensitive. For the T55 set, on-focus NPs show a steeper rise in F0 from t_n=8 to t_n=14 than broad-focus NPs, and then reach a similar F0 level as broad-focus NPs at t_n=18. This means that on-focus T55 NPs have a *domier* rise than the broad-focus variants. For the T25 set, broad-focus and on-focus NPs have comparable F0 at t_n=10, but the F0 of broad-focus NPs start rising at this same time point, while the F0 rise of on-focus NPs is delayed to t_n=14, where F0 also reaches a lower minimum than broad-focus NPs, and then rise to the same max F0 as broad-focus NPs at t_n=22. This means that on-focus T25 NPs set have a *scoopier* rise than the broad-focus variant. Situating these observations in TCoG terms, we expect that on-focus NPs will show *earlier* TCoG for T55 set, but *later* TCoG for T25 set, when compared to the respective broad-focus variants, effectively enhancing the timing difference between the two High targets. We now turn to verify these predictions.

Fig. 5 shows the TCoG value of NPs relative to the midpoint of the word-medial syllable in different focus conditions and tone sets, averaged across speakers. Values for broad-focus NPs are represented by the red vertical lines. The data is fitted to an LMEM following the same procedure reported in the previous section. The best model contains the interaction effect between FOCUS and TONESEQ. The expected changes in TCoG across focus conditions and tone sets are borne out. First, T25 NPs have significantly *later* TCoG than T55 NPs (β=122.0, t=6.89, p=.020), holding focus condition constant. This suggests that TCoG is useful in characterizing and differentiating T55 and T25 in Cantonese regardless of focus context. Concerning the T55 set, narrow- and contrastive-focus NPs have significantly earlier TCoG than the broad-focus NPs (CF: β=-9.42, t=-3.81, p=.0001; NF: β=-7.14, t=-2.89, p=.004).

Post-focus NPs, on the other hand, have comparable TCoG to broad-focus NPs. As for the T25 set, narrow- and contrastive-focus NPs have significantly later TCoG than broad-focus NPs (CF: β=21.24, t=8.11, p<.001; NF: β=15.81, t=6.06, p<.001). Interestingly, post-focus NPs in this set also have significantly earlier TCoG than broad-focus NPs (β=-6.64, t=-2.55, p=.0109). Finally, comparing the magnitude of focus-induced shift of TCoG, the T25 set exhibits a greater shift in TCoG than the T55 set (CF: 21.24 [T25] *vs.* 9.42 [T55]; NF: 15.81 [T25] *vs.* 7.14 [T55]).

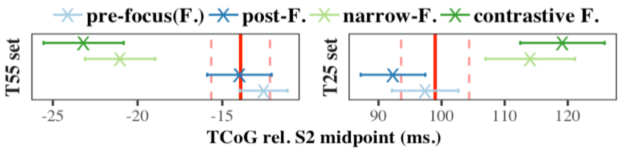

Figure 5: Mean TCoG relative to the midpoint of the NP-medial syllable. Red solid lines indicate group mean and SE of broad-focus NPs. Error bars and dashed lines indicate SE.

4. DISCUSSION

Previous studies argued that in Cantonese, focus is most consistently signaled by durational lengthening and intensity raising of the on-focus constituent. Pitch scaling is at best inconsistently modulated and mostly restricted to rising tones. PFC was also argued to be absent in Cantonese. The current study has gathered production data to demonstrate that our prior knowledge of Cantonese prosodic focus marking is at best incomplete. Based on data from 13 speakers, we find significant durational shortening and lowering of mean energy, max F0 & mean F0 of post-focus constituents relative to broad-focus ones. Our current dataset, however, is not best suited for explaining the discrepancy between our findings and previous ones. Whether it is due to task effects (e.g., adopting elicitation materials that promote *vs.* inhibit tonal coarticulatory variations) or a genuine change in prosodic focus marking strategies in Cantonese requires further investigations.

Turning to our specific research questions, this study presents evidence for the first time that Cantonese speakers modulate the contour shape of lexical tones as a response to information-structural differences. The direction and magnitude of focus-induced modulation of contour shape depend on tone type. In particular, relative to broad-focus target NPs, on-focus T55 shows a leftward shift of TCoG, but on-focus T25 shows a rightward shift. The observed tone-specific direction of change is in line with extant evidence presented in the segmental phonology literature that (sentential) stress leads to localized hyperarticulation (see e.g. [26], [27]). When the lexical tone is under focus, the High level tone (surrounded by two Low tones) is realized with a domier rise, hence more level-tone-like, while a High rising tone is realized with a later but faster (scoopier) rise, hence more rising-tone-like.

However, while hyperarticulated T55 flanked by Low targets adopting a dome-shaped rise seems intuitive, why should T25 be more scooped under focus? Recall that we observe that on-focus T25 reaches a lower F0 before they rise to the maximum F0 later on than T25 in other focus conditions (see also bottom left pane of Fig. 6 below). One potential motivation for the scooping of the T25 rise under focus could then be found in the assumption that this contour tone is best represented as a sequence of two levels, Low and High (see [28], cf. [29]). If this were true, then the scooped rise could serve to enhance simultaneously the later timing for the TCoG of the High tone, and lower scaling for the Low. Note however that the effect of focus on Low targets in our data is not consistent: The Low targets for the NP-medial T25 and NP-final T21 are lowered consistently, but that of the NP-initial T22 is not (see top panel of Fig. 6 below). This inconsistency may itself be serving to enhance contrast among the tones in the system, insofar as T21 is associated with lower F0 targets in general than T22. How Low targets are modulated under focus requires further investigation in the future.

The observation that at least some Low targets in our data (the Low of T25 and T21) are lowered under focus also casts doubts on the argument that pitch scaling is not modulated for narrow focus in Cantonese. To the extent that this lowering effect is true, pitch scaling is then actively modulated for focus purposes in Cantonese. Also related to pitch scaling is the well-documented correlation between contour shape and pitch scaling perception such that plateau-shaped peaks (e.g. green line of Fig. 1a) tend to sound higher than a sharp peak (e.g. black line of Fig. 1a) even if they have identical max. F0 at least to English listeners [30]–[32]. Assuming such findings are generalizable to Cantonese, on-focus T55 should then sound higher, even though max. F0 is not raised relative to broad-focus NPs. Whether the shape-scaling correlations are also true among Cantonese listeners, however, awaits experimental verification.

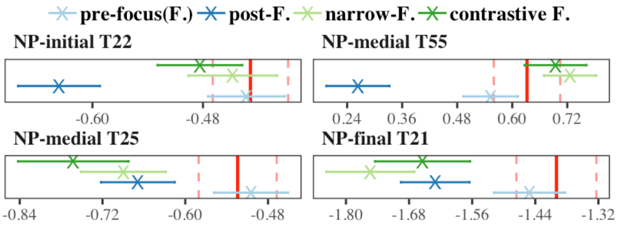

Figure 6: Mean Minimum F0 (z-score) of different tones in the target NPs. Red solid lines indicate values of respective tones in broad-focus condition. Error bars and dashed lines indicate SE.

5. REFERENCES

[1] D. Büring, “Towards a typology of focus realization,” in *Information structure: Theoretical, typological, and experimental perspectives*, M. Zimmermann and C. Féry, Eds. Oxford: Oxford University Press, 2010, pp. 177–205.

[2] F. Kügler and S. Calhoun, “Prosodic encoding of information structure: A typological perspective,” in *The Oxford Handbook of Language Prosody*, C. Gussenhoven and A. Chen, Eds. Oxford: Oxford University Press, 2020, pp. 454–467.

[3] W. E. Cooper, S. J. Eady, and P. R. Mueller, “Acoustical aspects of contrastive stress in question-answer contexts,” *J. Acoust. Soc. Am.*, vol. 77, no. 6, pp. 2142–2156, 1985.

[4] S. J. Eady and W. E. Cooper, “Speech intonation and focus location in matched statements and questions,” *J. Acoust. Soc. Am.*, vol. 80, no. 2, pp. 402–415, 1986.

[5] Y. Xu and C. X. Xu, “Phonetic realization of focus in English declarative intonation,” *J. Phon.*, vol. 33, no. 2, pp. 159–197, 2005.

[6] C. Féry and F. Kügler, “Pitch accent scaling on given, new and focused constituents in German,” *J. Phon.*, vol. 36, no. 4, pp. 680–703, 2008.

[7] Y. Xu, “Effects of tone and focus on the formation and alignment of f0 contours,” *J. Phon.*, vol. 27, no. 1, pp. 55–105, 1999.

[8] S. Ishihara, “Japanese focus prosody revisited: Freeing focus from prosodic phrasing,” *Lingua*, vol. 121, no. 13, pp. 1870–1889, 2011.

[9] R. S. Bauer, K.-H. Cheung, P.-M. Cheung, and L. Ng, “Acoustic correlates of focus-stress in Hong Kong Cantonese,” in *Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistics Society 2001*, S. Burusphat, Ed. Tempe, AZ: Arizona State University, 2004, pp. 29–49.

[10] W. L. Wu and L. Chung, “Post-focus compression in English-Cantonese bilingual speakers,” in *Proc. ICPhS XVII*, Hong Kong, 2011, pp. 148–151.

[11] W. L. Wu and Y. Xu, “Prosodic Focus in Hong Kong Cantonese without Post-focus Compression,” in *Proc. SP-2010*, Chicago, USA, 2010, Paper 040.

[12] W. Gu and T. Lee, “Effects of tone and emphatic focus on F0 contours of Cantonese speech: A comparison with standard Chinese,” *Chinese Journal of Phonetics*, vol. 2, pp. 133–147, 2009.

[13] V. C. H. Man, “Focus effects on Cantonese tones: An acoustic study,” in *Proc. SP-2002*, Aix-en-Provence, France, 2002, pp. 467–470.

[14] A. Arvaniti, D. R. Ladd, and I. Mennen, “Stability of tonal alignment: the case of Greek prenuclear accents,” *J. Phon.*, vol. 26, no. 1, pp. 3–25, 1998.

[15] J. Barnes, N. Veilleux, A. Brugos, and S. Shattuck-Hufnagel, “The effect of global F0 contour shape on the perception of tonal timing contrasts in American English intonation,” in *Proc. SP-2010*, Chicago, IL, USA, 2010, p. Paper 445.

[16] J. Barnes, N. Veilleux, A. Brugos, and S. Shattuck-Hufnagel, “Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology,” *Lab. Phonol.*, vol. 3, no. 2, pp. 337–383, 2012.

[17] J. Barnes, A. Brugos, N. Veilleux, and S. Shattuck-Hufnagel, “On (and off) ramps in intonational phonology: Rises, falls, and the Tonal Center of Gravity,” *J. Phon.*, vol. 85, 101020, 2021.

[18] M. D’Imperio, “The role of perception in defining tonal targets and their alignment,” PhD dissertation, The Ohio State University, 2000.

[19] D. R. Ladd, D. Faulkner, H. Faulkner, and A. Schepman, “Constant ‘segmental anchoring’ of F0 movements under changes in speech rate,” *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1543–1554, 1999.

[20] D. R. Ladd, *Intonational Phonology*, 2nd ed. Oxford: Oxford University Press, 2008.

[21] O. Niebuhr, M. D’Imperio, B. G. Fivela, and F. Cangemi, “Are there ‘shapers’ and ‘aligners’? Individual differences in signalling pitch accent category,” in *Proc. ICPhS XVII*, Hong Kong, 2011, pp. 120–123.

[22] J. ’t Hart, “F0 stylization in speech: straight lines versus parabolas,” *J. Acoust. Soc. Am.*, vol. 90, no. 6, pp. 3368–3370, 1991.

[23] P. Boersma and D. Weenink, *Praat: doing phonetics by computer*. 2020. [Online]. Available: http://www.praat.org/

[24] Y.-L. Shue, P. Keating, C. Vicenik, and K. Yu, “Voicesauce: A program for voice analysis,” in *Proc. ICPhS XVII*, Hong Kong, 2011, pp. 1846–1849.

[25] D. Talkin, “REAPER: Robust and epoch and pitch estimator.” 2014. [Online]. Available: https://github.com/google/REAPER

[26] K. J. de Jong, “The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation,” *J. Acoust. Soc. Am.*, vol. 97, no. 1, pp. 491–504, 1995.

[27] K. de Jong, M. E. Beckman, and J. Edwards, “The interplay between prosodic structure and coarticulation,” *Lang. Speech*, vol. 36, no. 2–3, pp. 197–212, 1993.

[28] J. L. Lee, “The Representation of Contour Tones in Cantonese,” in *Proceedings of the Annual Meeting of the Berkeley Linguistics Society 38*, 2012, pp. 272–286.

[29] M. Barrie, “Contour tones and contrast in Chinese languages,” *J. East Asian Ling.*, vol. 16, no. 4, pp. 337–362, 2007.

[30] J. Barnes, A. Brugos, N. Veilleux, and S. Shattuck-Hufnagel, “Voiceless intervals and perceptual completion in F0 contours: Evidence from scaling perception in American English,” in *Proc. ICPhS XVII*, Hong Kong, 2011, pp. 108–111.

[31] J. Barnes, A. Brugos, N. Veilleux, and S. Shattuck-Hufnagel, “Segmental influences on the perception of pitch accent scaling in English,” in *Proc. SP-2014*, Dublin, Ireland, May 2014, pp. 1125–1129.

[32] R.-A. Knight, “The shape of nuclear falls and their effect on the perception of pitch and prominence: peaks vs. plateaux,” *Lang. Speech*, vol. 51, no. Pt 3, pp. 223–244, 2008.

[1] Cantonese has six lexical tones: high level (T55), mid level (T33), low level (T22), high rising (T25), low rising (T23), low falling (T21). The short-hand notations are based on Chao’s tone numerals, with *1* referring to a speaker's lowest pitch and *5* highest.

[2] More complex random effect structures were not considered because such moves resulted in model singularity.