

Effects of segmental composition on f0 peak alignment of Cantonese rising tones

Ricky KW Chan (kwrc2@cam.ac.uk) and Yang Li (yl492@cam.ac.uk)

Phonetics Laboratory, University of Cambridge

Introduction. The study of how fundamental frequency (f0) aligns with segments has important implications for not only description of tonal and intonational systems, but also theoretical modelling and synthesis of prosody. For example, Xu (1998) found that the f0 peak of the rising tone in Mandarin consistently occurs after the offset of the tone-carrying syllable regardless of speaking rate and whether the syllable has a final nasal. This has been taken as evidence that tone is fully syllable-synchronised, with the peak delay attributed to inertia. This tone-syllable synchrony serves as a foundational assumption of the Time Structure model of the syllable (Xu & Liu, 2006) and PENTA model (Xu, 2005). However, it remains unclear whether this synchrony can be extended to other tone languages. Here we report an experiment on the effects of segmental composition on f0 peak alignment of the two rising tones (i.e. T2[25] and T5[23]) in Cantonese.

Method. Meaningful monosyllables /ji:/, /ŋa:/, /ji:m/, and /si:/ were concatenated to form three types of disyllabic non-words—/ji: ŋa:/, /ji:m ji:/, and /ji: si:/. 9 native speakers of Hong Kong Cantonese (5m and 4f, aged 19-28) read aloud these disyllables in a carrier sentence “我講xx俾你聽” (“I say xx to you”). To control for speaking rate, regular beats were played at an interval of 3s and subjects were instructed to read aloud each sentence between two beats. 20 f0 values of the tones were extracted using the VoiceSauce (Shue, 2010) with the STRAIGHT algorithm (Kawahara et al., 1998).

Results. Figures 1 to 3 show the f0 contour of the tones in three types of disyllabic non-words, in which the first syllable carries one of the two rising tones. Except when followed by the high level tone, the f0 peaks of the two rising tones predominantly occur on the first syllable for /ji:m ji:/, but either close to the syllable boundary or on the second syllable for /ji: ŋa:/. These patterns are highly consistent across speakers. The f0 peak of the rising tones in /ji: si:/ occurs before the syllable boundary, but the pre-boundary fall in f0 could be attributable to spread glottis in anticipation of the voiceless fricative /s/. It appears that segmental composition affects f0 peak alignment of lexical tones in a way reminiscent of intonational tones (D’Imperio, 2000). No segmental anchoring of f0 peak was observed, and its absence cannot be attributed to time pressure differences on the first syllable (Ladd et al., 1999; Schepman et al., 2006), as the two rising tones have almost identical peak locations in all three syllable compositions. Moreover, unlike Mandarin (Xu, 1998), the peak of the Cantonese rising tones may occur well before the syllable boundary (e.g. /jim: ji/), and the early peaks cannot be the product of inertia and must be pre-specified. This weighs against tone-syllable synchrony as a universal principle, a crucial assumption of the Time-Structure model and PENTA model. Finally, the lack of consistent alignment of f0 peak casts doubt on the usefulness of classifying tone languages based on the relative degree of carry-over and anticipatory coarticulation, e.g. the same tonal sequence in Cantonese can show more anticipatory coarticulation in /ji:m ji/ than in /ji: ŋa:/, suggesting that future studies on coarticulation need to take segmental compositions into account.

References

- D’Imperio, M. (2000). *The role of perception in defining tonal targets and their alignment*. Doctoral dissertation, Ohio State University.
- Kawahara, H., de Cheveign, A., and Patterson, R. D. (1998). An instantaneous-frequency-based pitch extraction method for high quality speech transformation: Revised TEMPO in the STRAIGHT-suite. In *Proc. ICSLP’98*, Sydney, Australia.
- Ladd, D. R., Faulkner, D., Faulkner, H., & Schepman, A. (1999). Constant “segmental anchoring” of F0 movements under changes in speech rate. *The Journal of the Acoustical Society of America* 106: 1543–1554.
- Schepman, A., Lickley, R., & Ladd, D. R. (2006). Effects of vowel length and “right context” on the alignment of Dutch nuclear accents. *Journal of Phonetics* 34: 1–28.
- Shue, Y.-L. (2010). *The voice source in speech production: Data, analysis and models*. Doctoral dissertation, UCLA.
- Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55: 179–203.

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46: 220-251.
 Xu, Y. and Liu, F. (2006). Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Italian Journal of Linguistics* 18: 125-159.

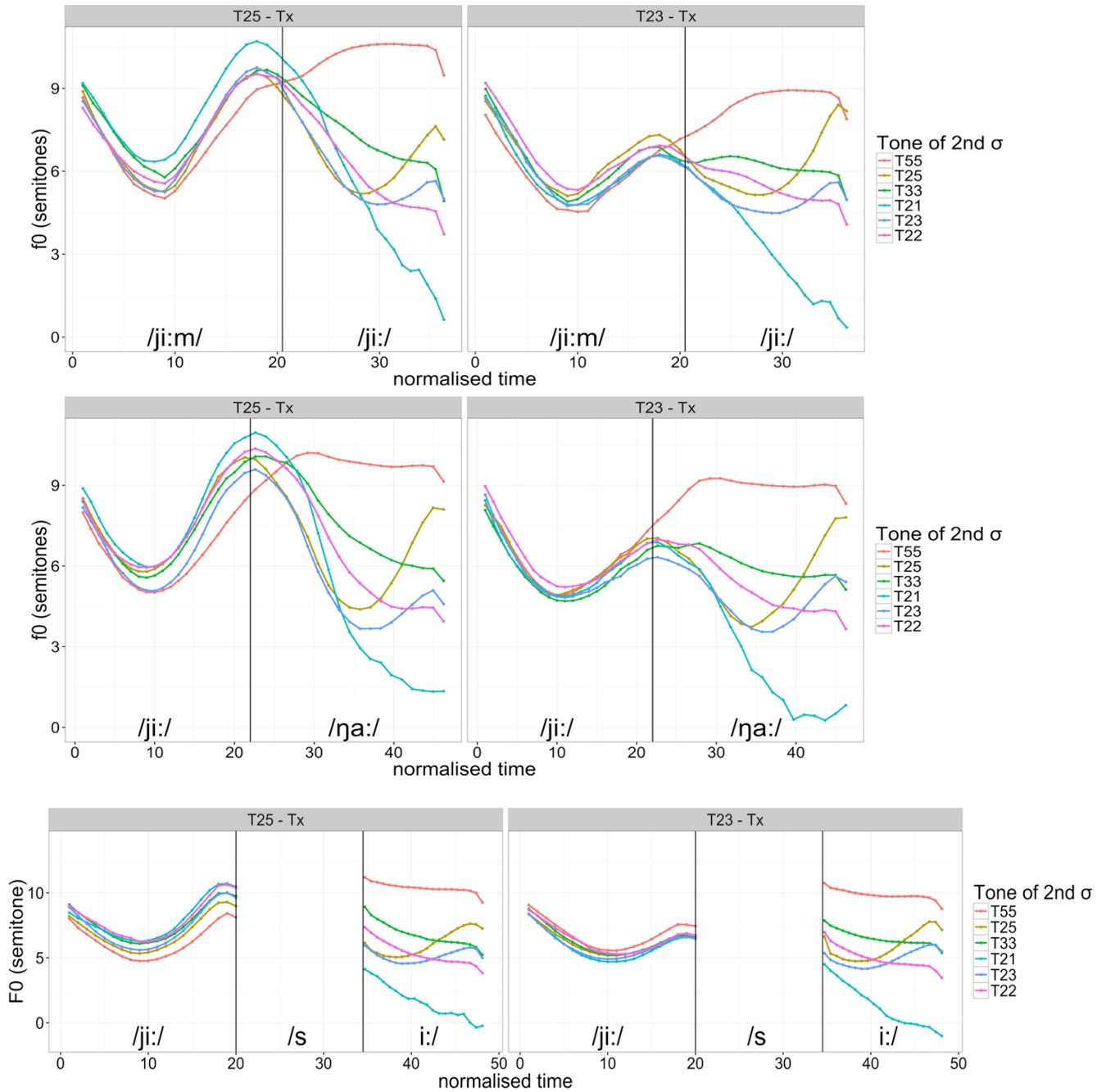


Figure 1 to 3. f0 contours of T2[25]-Tx (left panels) and T5[23]-Tx (right panels) disyllabic non-words /ji:m ji:/, /ji: ɲa:/, and /ji: si:/ respectively. X-axis depicts the normalised durational ratio of the segments; y-axis represents the raw f0 values in semitones. Vertical lines represent the location of the syllable boundary (in Figure 1 and 2) and segmental boundary (in Figure 3). The f0 contours of /s/ in Figure 3 are removed as the estimate is unreliable.