12th ISSP

Dec. 14-18, 2020

Acoustic correlates of question-final Cantonese tones

Una Y. Chow

una.chow@ubc.ca

Introduction

Cantonese tones

- High-, mid-, low-level (T1, T3, T6); high-rise, low-rise (T2, T5); low-fall (T4) [1]

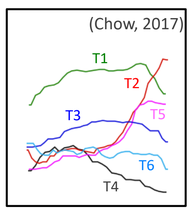

Figure 1. Cantonese tones

Previous study (Khouw & Ciocca, 2007) [2]

- Discriminant analysis on f0 of 8 consecutive sections of the vocalic portion of 24 monosyllabic words, read by 10 Cantonese speakers
- Magnitude of f0 change over the 6th and 7th sections of the tone was important in identifying the tones.

Research questions

A tone likely differs in shape in continuous speech than in isolated words due to intonational effects.

1. To what extent can **f0 height** and **magnitude of f0 change** serve as acoustic cues to Cantonese tones on final syllables of declarative questions?
2. At which portion of the tone can these two acoustic properties best separate the tones?

Cantonese data

Dataset

- Extracted from a corpus [3] developed for analyses of Cantonese intonation:
  - 10 native **Hong Kong Cantonese speakers** (5 male, 5 female; age 21-26 years)
  - 5 blocks of 4 dialogues
    - each block contains the same target final syllable segments (e.g., *si*)
    - each dialogue contains a declarative question ending in one of six Cantonese tones
    - each target tone is different within a dialogue (e.g., Figure 2)
- **Four tonal groups:** H (T1), M/L (T3, T6), R (T2, T5), F (T4)
  - For the current analysis, T3 was combined with T6 and T5 with T2 because the paired tones appear to be merging for some speakers [4, 5] and there were few tokens of them.
- **400 tokens** of question-final syllables:
  - 10 speakers x 5 syllables x 4 tonal groups x 2 readings

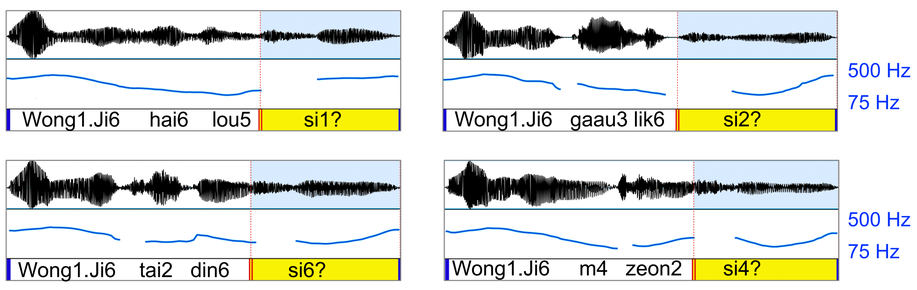

Figure 2. F0 contours of question-final syllables ending with *si1* (H), *si2* (R), *si6* (L), and *si4* (F); female speaker.

Methods

F0 extraction (using Praat [6])

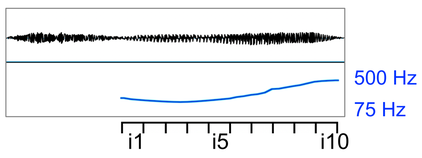

Figure 3. F0 extraction from the periodic rhyme of the final syllable (e.g., *si2*) at 10 equal intervals.

1. **F0 height** (mean f0, in semitone) at each interval (i1, i2, i3, ..., i10)
2. **Magnitude of f0 change** (absolute mean f0 difference) between intervals (i1-i2, i2-i3, i3-i4, ..., i9-i10)

Random forest classification

- Conditional random forest (*cforest*) from the `partykit` package in R [7]
- A classification method that generates a number of recursively binary-splitting trees (= 1000, in this study)
- *Cforest*'s overall prediction is based on the aggregate results of the individual trees.
- 2-fold cross-validation: tokens from the first and second readings alternated as training and test data

Results

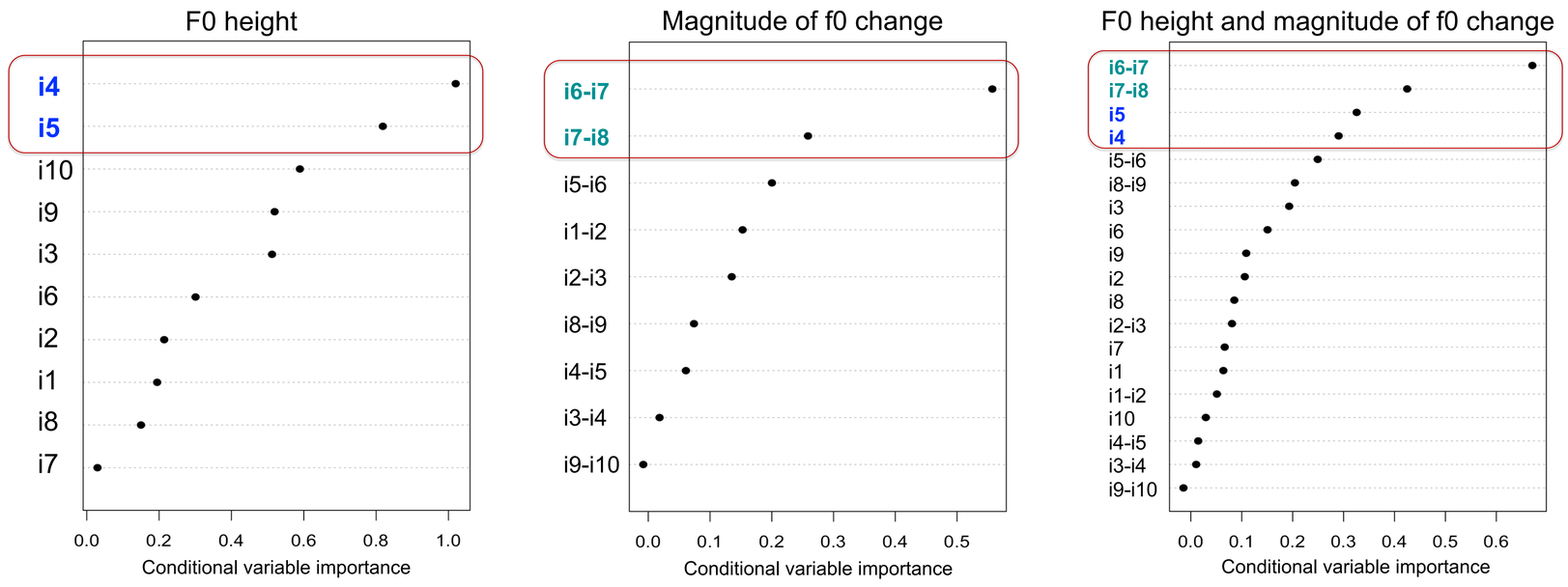

Figure 4. Relative importance of the different sections of the tone in separating it from the other tones, based on f0 height (left chart), magnitude of f0 change (middle chart), and both evaluated together (right chart), estimated by *cforest*. Rank orders from the first 2-fold cross-validation tests are shown.

Table 1. Sections of the tone that are ranked the most important by *cforest* in predicting the tones

	Cross-validation test 1	Cross-validation test 2
F0 height	**i4**, i5, i10	**i4**, i10, i5
Magnitude of f0 change	**i6-i7**, i7-i8	**i6-i7**, i7-i8
Both properties	**i6-i7**, i7-i8, **i5, i4**	**i6-i7**, i7-i8, i8-i9, **i4, i5**

Table 2. Balanced accuracy (mean of sensitivy and specificity): average of all four tonal groups and both tests

	Balanced accuracy
F0 height	0.7467
Magnitude of f0 change	0.7384
Both properties	0.7684

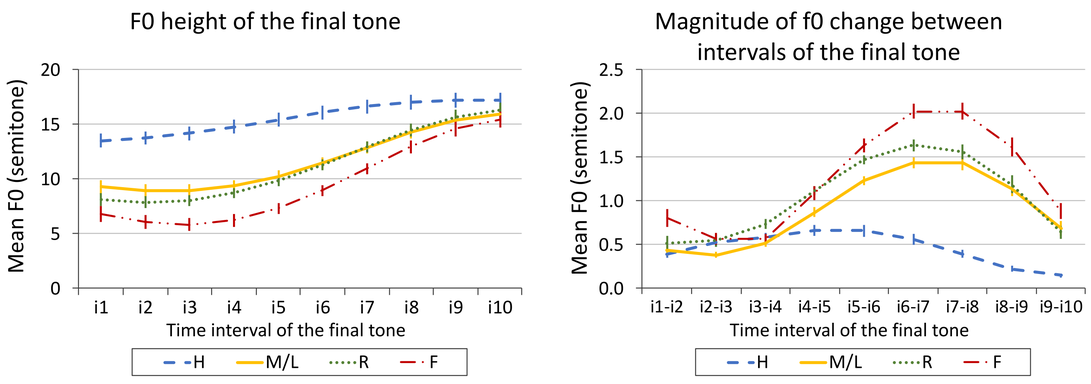

Figure 5. Mean f0 values at each of the 10 intervals of the final tone (left) and magnitude of f0 change between intervals of the final tone (right), by tonal group.

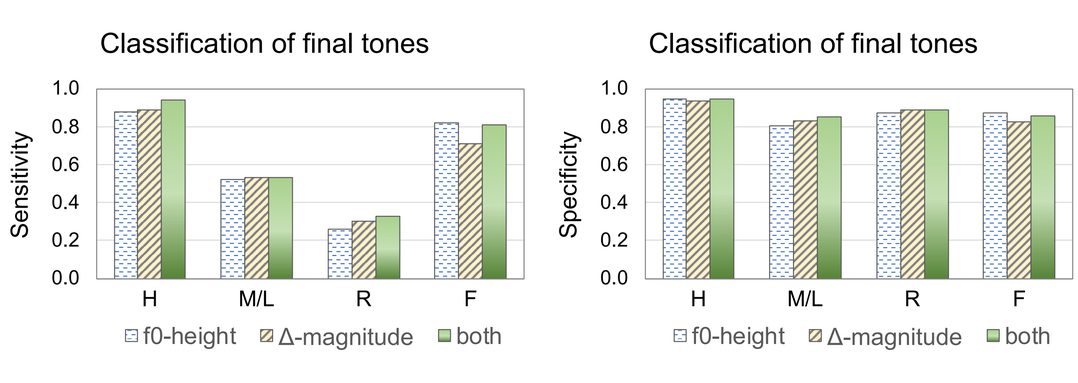

Figure 6. Sensitivity and specificity values of *cforest*'s correct identification and correct rejection, respectively, of the final tones, averaged between both cross-validation tests.

Discussion

F0 as acoustic cue to Cantonese tones

- F0 height and magnitude of f0 change both performed well as acoustic predictors of the H and F tones but poorly of the M/L and R tones (Figure 6, left).
  - The H tone is produced with a much higher f0 than the other tones so it is clearly separate from them (Figure 5).
  - The F tone has a dip before the final rise, which makes it more distinct as well.
  - The M/L and R tones, in canonical form ([33]/[22] and [25]/[23], respectively), are acoustically distinct from each other by the f0 height or f0 change at the second part of the tone. This distinction is neutralized by the final question rise. See Figure 5.

Most distinctive section of the tone

- **For f0 height:** the 4th interval of the tone was estimated to be the most important (Figure 4). This is approximately where the M/L/R/F tones start ascending and converging toward the target, high question tone (Figure 5).
- **For magnitude of f0 change:** between the 6th and 7th intervals was estimated to be the most important. This is consistent with Khouw & Ciocca's (2007) result.
- When both acoustic properties were included in a single model, the magnitude of f0 change betweeen the 6th and 7th intervals outranked the f0 height at the 4th interval, suggesting that the former provides a more robust cue. The balanced accuracy score is slightly higher with the addition of f0 height (Table 2) so the latter also played a role.

Conclusion

- F0 analysis of question-final tones indicates that the magnitude of f0 change between the 6th and 7th intervals can serve as a reasonably well cue for tonal identification but for certain tones (H and F) only. The neutralization of the later part of the tone by the high question rise calls for non-f0 acoustic cues.

Future directions

- Analyze voice quality across different sections of the final tone. Question-final tones may be less creaky than statement-final tones due to final rise in pitch.
- Conduct a perception study to see which portion(s) of the tone listeners perceive to be important in identifying the question-final tones

References

[1] Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese Phonology*. Mouton de Gruyter.

[2] Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics*, *35*, 104–117.

[3] Chow, U. Y. (2017). *An exemplar-based model of intonation perception of statements and questions in English, Cantonese, and Mandarin* (Master's thesis, University of Calgary, Canada).

[4] Mok, P. P., Zuo, D., & Wong, P. W. (2013). Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change*, *25*(3), 341–370.

[5] Whitehill, T. L., Ciocca, V., & Chow, T. Y. (2000). Acoustic analysis of lexical tone contrast in dysarthria. *Journal of Medical Speech—Language Pathology*, *8*, 337–344.

[6] Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer* (Version 6.1.05). Retrieved from http://www.praat.org

[7] R Core Team. (2019). R: A language and environment for statistical computing (Version 3.6.1) [Computer application]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved July 18, 2019, from https://www.R-project.org/

Acknowledgements:

Thanks to the reviewers of ISSP 2020 for their helpful comments on the abstract.

This research, including data collection, was supported by the Social Sciences and Humanities Research Council of Canada via a Canada Graduate Scholarship.

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada