Quantitative analysis of intonation patterns in statements and questions in Cantonese

Joan K-Y Ma, Valter Ciocca & Tara L Whitehill

Division of Speech and Hearing Sciences University of Hong Kong

{joanma; vciocca; tara}@hkusua.hku.hk

Abstract

The aim of this study was to investigate intonation patterns in Cantonese using a quantitative approach. The commandresponse model was employed to explore the differences between intonations, and the effects of lexical tone on fundamental frequency (F0) contours of intonation. Two intonation types (statements and questions), with six tonal contrasts embedded at the final position of the utterance, were collected from twelve native Cantonese speakers (six males and six females). Results showed that F0 in questions was raised for the entire utterance, which was mainly associated with baseline frequency changes. An additional rise in the F0 contour of questions, which was represented by tone command changes, was observed from the midpoint of the sentence onward. An additional positive boundary tone command occurred towards the end of the final syllable of questions, which denoted the final-rise in F0 in questions. A lengthened duration of the tone command towards the end of the utterance in questions was also observed. The amplitude of the final-rise in the contours of questions was affected by the tone of the final syllable, with significantly higher amplitude noted for the boundary tone command of tones 25 and 21.

1. Introduction

Intonation is a universal feature shared by languages of different origins. It plays an important role in conveying both linguistic and paralinguistic meaning in communication. Variations of fundamental frequency (F0) at sentence level are used to signify intonation. The co-occurrence of lexical tone and intonation in tone languages such as Cantonese poses an interesting research question concerning how the global F0 movement (i.e., intonation) can be separated from local F0 changes (i.e., tone). The overall F0 pattern observed is considered a superimposition of tone and intonation values [1 & 3]. Chao [3] viewed the interplay between tone and intonation as "small ripples riding in large waves" (p.39). He suggested that the actual value of F0 would be the algebraic sum of tone and intonation by synchronic superposition.

There have been few investigations of intonation patterns in Cantonese. Chang [2] used intonation group as a prosodic unit in analysing the intonation contours, in terms of pitch height and pitch span, in Cantonese connected speech. However, this method of analysis does not provide quantitative measures for intonation events, such as the magnitude of increase in F0 height, which would be necessary for statistical comparison between intonation events and acoustic data. Gu, Hirose and Fujisaki [9, 10 & 11] adopted a quantitative model to analyse intonation in Cantonese. Using the commandresponse model [8], the F0 contour of intonation on a logarithmic scale was described as the sum of a phrase component, a tone component and a baseline level. This model allows direct comparison of parameters, such as the amplitude of final-rise in questions across different tones [e.g. 11]. However, most of the works of Fujisaki and colleagues has focused on optimization of different model parameters. Further research is needed to investigate the relationship between the model parameters and acoustic or linguistic variables.

The intonation patterns of Cantonese are similar to those of the many other languages [6]. Generally speaking, statements have a falling F0 contour, while questions have a rising F0 contour. In this study, statements and intonation questions (i.e., a declarative sentence turned into question by superimposing a rising contour) are investigated. Vance [17] showed evidence of declination in statements in Cantonese. Ma, Ciocca and Whitehill [14] showed that the same tone has a higher relative F0 level in the middle of a sentence than at the final position in statements. As for intonation questions, Chang [2] showed that the general downtrend was still observed, followed by a rising contour which affects only the last syllable of the utterance. However, acoustic analysis was performed on only one sentence in her experiment. Other studies [6 & 14] supported the claim that tone contours at the final position of an utterance would be modified to rising by question regardless of the canonical form of the tone. Although the overall F0 contour of intonation questions was not examined, Ma, Ciocca and Whitehill [14] also found that tones produced in questions have a higher F0 level than in statements. This is consistent with Ho's [13] observation that questions have a higher F0 level across the entire sentence in Mandarin.

In exploring the interaction between tone and intonation, Fry [7] viewed tone as subordinate to intonation and likely to be modified by intonation. However, Yuan [18] showed that tones at the final syllable of an utterance affect intonation identification in Mandarin, such as questions ending with a rising tone were harder to identify than questions ending with a falling tone. He hypothesized that statement is a default choice in listeners' perception, and questions would only be identified when the intonation marking (e.g. final-rise in F0) is clearly presented. Therefore, rising tones at the final position may confuse the listeners as the final-rise marking is ambiguous. There are six contrastive tones in Cantonese: high-level (55), high-rising (25), mid-level (33), low-falling (21), low-rising (23) and low-level (22). Although it was shown that questions modify the F0 contour of tones at the final position [15], the amplitude of F0 changes may differ among tones. Yuan [18] suggested that tone changes have to be considered in using intonation model in tone languages.

The purpose of this study is to investigate intonation patterns in Cantonese, by using the command-response model, and to relate the findings to previous acoustic and perceptual analyses. The effect of tone on intonation production will also be examined.

2. Methods

2.1. Speakers

Twelve native Cantonese speakers were recruited (six males and six females, age 19 to 25). They were all undergraduates or graduates from local universities. Cantonese was reported to be the native language for all speakers, while English was the second language. All speakers had no reported history of speech problems and all passed a hearing screening (\leq 20 dBHL at 250, 500, 1000, 2000 and 4000 Hz).

2.2. Speech materials

Two intonation patterns were studied: statement and intonation question. The carrier phrase /lei₅₅ ko₃₃ tsi₂₂ hei₂₂ <u>X</u>/ (This word is <u>X</u>), which can be produced as either a question or a statement, was used in this experiment. Three sets of words, derived from the syllables /si/, /ji/ and /jeu/, were embedded in the above sentence context. Each set consisted of six words that differed only in tone. With two intonations and 18 words, a total of 36 stimuli were collected from each speaker.

2.3. Data collection

Data collection was carried out in a sound-attenuated room (IAC sound-proof booth), with a Sony TCD-D3 DAT recorder and a Bruel & Kjaer (4003) low-noise unidirectional microphone. A 10 cm mouth-to-microphone distance was maintained during recordings. The dialogues were presented visually on the screen of a G4 Apple Macintosh running a Hyper-Card (AppleTM) programme. The sequence of presentation of the 36 stimuli was randomized across subjects. In each trial, one of the dialogues was presented on the screen. In order to make the speech sample as natural as possible, the subject was engaged in a guided conversation with the first author.

After recording, each token was low-pass filtered at 22k Hz, digitized at sampling rate of 44.1k Hz and recorded onto the computer as separate file using a DigiDesign Audiomedia II DSP card on a Apple PowerMacintosh 7100 computer.

2.4. Data analysis

The F0 pattern of the stimuli was analysed according to the command-response model [8]. In this model, the F0 pattern of a sentence is considered as a linear superposition of the phrase component (denoting the global F0 changes, i.e., intonation) and the tone component (denoting the local F0 changes, e.g. tone) on a baseline frequency. The formulas for the model can be found in [8].

According to [8], the base frequency (F_b) is the F0 baseline of a sentence in the absence of other components. The phrase component is produced by the impulse responses (phrase commands) of the phrase control mechanism. Each phrase command is defined by the magnitude of the command (A_p) and the onset time (T₀). There is also a time constant (α) for the phrase control mechanism. Tone component is a response of the step-wise input signals (tone commands) of the tone control mechanism. The tone commands of both positive and negative polarities are defined by amplitude (A_a), onset time (T₁) and offset time (T₂). There is a time constant (β) for the tone control mechanism, and also a ceiling level (γ) of the tone command. Both mechanisms are assumed to be critically-damped second-order linear systems.

In the current study, the parameters F_b , A_{pi} and T_{0i} of the i^{th} phrase command, and A_{tj} , T_{1j} and T_{2j} of the j^{th} accent

command were used to quantify the F0 pattern. In addition, the duration of each tone command (D_{ti}) was obtained by calculating the difference in onset and offset time $(T_{2i} - T_{1i})$. FujiParaEditor [16] was used for analysis. The F0 pattern was extracted by an autocorrelation algorithm. Automatic calculation was used for initial analysis, followed by manual adjustment using the Analysis-by-Synthesis. For each stimulus, the F_b was set as the lowest F0 value within the utterance. The number of phrase and tone commands was determined by linguistic constraints such as syllable boundary, and the magnitude and timing of each command was approximated to generate the closest F0 contour resembling the actual utterance. A set of tone command patterns for Cantonese has been proposed [9]. They are positive for tone 55, negative followed by positive for tone 25, zero for tone 33, negative for tone 21, negative followed by zero for tone 23 and negative for tone 22. Previous research has shown that the final-rise in F0 in questions can be generated by an addition of tone command (i.e., boundary tone command) [11]. Parameters α and β are assumed to be constant within a stimuli, and were set as the default value of the program (2 and 20 respectively). The parameter γ was set as 0.9 in this experiment.

3. Results

Figure 1 shows an example of the analysis using the commandresponse model. The mean value for each parameter (F_b , A_{pi} , T_{0i} , A_{tj} , and D_{tj}) of each tone at each intonation was calculated for each speaker. Three-way ANOVAs were used to analyse the data for each parameter separately, with gender as the between subject factor, and intonation and embedded tone as the within subject factors.

3.1. Base frequency

Statistical analysis showed significant main effects for gender, intonation and embedded tones. The F_b of female speakers was significantly higher than that of the male speakers [F (1, 10) = 8.92, p < 0.05]. Questions had a significantly higher F_b than statements [F (1, 10) = 31.52, p < 0.05]. There were also significant differences in F_b between the embedded tones [F (5, 50) = 26.62, p < 0.05]. Tone 33 embedded sentences had the highest F_b value, with significant difference with tones 25, 21, 23 and 22 (Tukey HSD test, p < 0.05 for all) but not with tone 55 (Tukey HSD test, p > 0.05). Tone 21 embedded sentences had an F_b value significantly lower than the other five tones (Tukey HSD, p < 0.05 for all).

3.2. Phrase component

Two phrase commands, as derived from the prosodic structure of the speaker's production, were used to model the phrase component for the stimuli in the present experiment. A fourway ANOVA was used to compare the magnitude of the phrase commands, as the difference between the two phrase commands was also considered. The results showed a significant main effect for embedded tones [F(5, 50) = 23.35, p]< 0.05]. Post-hoc analysis showed that the A_p of tone 21 embedded sentences was significantly higher than the sentences embedded with the other five tones (Tukey HSD test, p < 0.05for all). A significant difference was also noted in the magnitude of the two phrase commands, with Ap1 significantly higher than A_{p2} [F (1, 10) = 9.22, p < 0.05]. The onset time (T_0) of each of the two phrase commands were compared separately. There was no significant difference in T_0 across gender, intonation and embedded tones (p > 0.05 for all).



Figure 1. Example of intonation analysis by commandresponse model of question intonation. The four panels from top to bottom represent the waveform, the F0 contour as a function of time, the tone and the phrase components.

3.3. Tone component

The number of tone command was used to model the local F0 changes of tones in each syllable depends on the tone of that syllable, as stated earlier. In addition, a positive boundary tone command was added towards the end of the utterance for questions. Statistical analysis comparing the amplitude of each tone command showed that, in the carrier phrase /lei55 ko33 tsi22 hei22/, a significant main effect for intonation was noted for words 3 (/tsi₂₂/) and 4 (/hei₂₂/) [F(1, 10) = 23.97, p < 1000.05 for A_{t3}; and F (1, 10) = 6.41, p < 0.05 for A_{t4}], with the amplitude of the tone command higher in questions. A significant main effect of the embedded tone was also noted for both A_{t3} and A_{t4} [*F* (5, 50) = 5.23, *p* < 0.05 for A_{t3} ; and *F* (5, 50) = 11.99, p < 0.05 for A_{t4}]. Post-hoc analysis revealed that tone 33 embedded sentences have significantly lower At3 and A_{t4} than sentences embedded with the other five tones (Tukey HSD test, p < 0.05 for all). In analysing the amplitude of the embedded tone (At5), a two-way ANOVA, with the factors gender and intonation, was performed for each tone. Results showed a significant main effect of intonation for tone 55 [F(1, 10) = 54.02, p < 0.05], tone 25 [F(1, 10) = 25.82, p< 0.05], tone 21 [F (1, 10) = 12.87, p < 0.05], and tone 23 [F (1, 10) = 205.42, p < 0.05, but not for tone 22 [F (1, 10) = 2.10, p > 0.05]. A two-way ANOVA was used to compare the At6 of the boundary tone command in questions across gender and tones. Female speakers produced significantly higher Att6 than male speakers [F(1, 10) = 5.74, p < 0.05]. A significant difference was also noted for the main effect of embedded tone [F (5, 50) = 17.66, p < 0.05]. Post hoc analysis showed that tones 25 and 21 had significantly higher Att (Tukey HSD test, p < 0.05 for all), except for the difference between tone 21 and tone 23 (Tukey HSD test, p > 0.05). The other four tones (tones 55, 33, 22 and 22) showed similar Att6 (Tukey HSD test, p > 0.05 for all), except that the difference between tones 23 and 33 was significant (Tukey HSD test, p < 0.05).

As mentioned earlier, the duration of each tone component was calculated by the difference between T_{1j} and T_{2j} . The total duration of two tone commands was used instead when two commands were needed in the modelling of

some tones or in the case of questions in the final syllable. Statistical analysis showed a significant main effect for gender for tone command 1 [F(1, 10) = 9.44, p < 0.05], with female speakers producing a longer duration for tone command 1. A significant main effect for embedded tone was observed in tone commands 1 [F(5, 50) = 6.11, p < 0.05] and 4 [F(5, 50)= 14.48, p < 0.05]. Post-hoc analysis showed that, in both tone commands 1 and 4, the Dt of tones 55 and 33 embedded sentences were significantly shorter than in those embedded with the other four tones (Tukey HSD test, p < 0.05 for all), except for the difference between D_{t1} of tone 25 and tone 33 embedded sentences (Tukey HSD test, p > 0.05). Statistical analysis of the duration of the embedded tone showed a significant main effect of intonation [F (1, 10) = 31.88, p < p0.05] and tone [F (5, 50) = 10.55, p < 0.05], where tone command for questions has a longer duration than for statements.

4. Discussion

This experiment was to analyse the intonation patterns in Cantonese using a quantitative model. The phrase component showed an overall pattern of declination in the current experiment. The phrase component was modelled by two phrase commands, with A_{p2} significantly lower than A_{p1} . Fujisaki and Hirose [8] described the occurrence of the second phrase command as re-phrasing, with the second phrase common the first. Therefore, the overall phrase component can be described as declination for both intonation questions and statements. This is consistent with previous reports on the acoustic data of F0 patterns of statements [14, 17] and intonation questions [2] in Cantonese.

Although declination was observed in the global pattern for both questions and statements, speakers are known to mark question intonations by both raising the global F0 level of the whole sentence and a final-rise in F0 in English [4], Mandarin [13] and Cantonese [2, 14]. The results of the current experiment showed statistically significant differences between questions and statements in F_b. As F_b was set as the lowest F0 value in each sentence in this experiment, the difference in F_b between questions and statements suggests that, in marking the global increase in F0 level in questions, speakers raised the F0 level by a certain value throughout the entire sentence. Differences between the two intonations were also noted in the A_t for words 3, 4 and 5 of all the stimuli, except for Atts of tone 22 embedded sentences. This shows that in addition to the global marking of intonation, local changes, as represented by an increase in the amplitude of tone commands, also occurred. Although the rise in F0 was most prominent at the final position, subtle but statistically significant differences were also noted from the mid position of the sentence onwards for marking questions. However, no significant difference in Att5 was noted between the two intonations for tone 22 placed at the final position. This is hypothesized to be related to the relative distance of tones. As tones 33 and 22 have a closer relative distance together than tones 55 and 33, the relatively restricted tonal space of tone 22 might have placed a constraint for the increase in tone level when produced as a question.

In modelling the final-rise in questions, a positive boundary tone command was used to approximate the F0 changes. Although the involvement of a mechanism other than tone component has been speculated [8], the addition of the boundary tone command corresponded well to the acoustic analysis, as a rising F0 contour was observed for all tones at the final position, regardless of the canonical form, in Cantonese [14]. Additionally, previous perceptual analysis showed that identity of the tones at the final position of questions were modified, as listeners showed misidentification of tones 33, 21, 23 and 22 as tone 25 [15].

An additional finding in the current experiment was that questions have a significantly longer duration for the final tone command than the same word produced in statements. This was consistent with the durational measure from our acoustic analysis, where questions (mean = 0.321, SD = 0.04) showed a longer duration than statements (mean = 0.28, SD = 0.05). Eady and Cooper [4] had similar findings in analyzing the syllable duration of English intonation. They hypothesized that the durational difference is related to the relative semantic prominence of the sentence-final syllable in questions in comparison to statements.

The amplitudes of the boundary tone commands differed among the six tones, showing that the tone at utterance final position has an impact on the F0 contour of the utterance. Tones 25 and 21 had a significantly higher A_t than the other four tones for the boundary tone command, possibly due to the canonical structure of Cantonese tones. Tone 21 has a falling contour [6], and although it starts at about the same height as tone 22, it falls to the lowest F0 level among the six tones [14]. As a result, a larger degree of amplitude change would be expected for the boundary tone command in order to have a relative F0 level comparable with the other tones. Tone 25 has a rising contour in the canonical form [6], and the rise in F0 was presumably accentuated when placed at the final position of questions in order for the question marking to be distinctive, resulting in a greater A_t value for the boundary tone command.

Using the command-response model, both male and female were found to have similar phrase components and tone components across intonations. Although female speakers have a higher F0 than male speakers, the difference in F0 was equalized by setting the F_b to the lowest value of the utterance. Differences in other parameters were noted between the two genders. Female speakers produced the first tone command with a lengthened duration. Significantly higher amplitude was also noted for the boundary tone command in female speakers. Female speakers are known to have wider F0 range than male speakers [12], which may explain the difference in amplitude of the boundary tone command. Fitzsimons, Sheahan and Staunton [5] suggested that F0 range is a function of speech rate, where increased speech rate places a restriction on the potential F0 range. Therefore, the lengthened syllable in female speakers may result in a slower speech rate, and hence greater variability in the amplitude of the boundary tone command.

5. Conclusion

The present study examined two intonation patterns in Cantonese by quantitative approach. Both statements and questions showed declination in F0 contour. However, questions were marked by an increase in the overall F0 level, which is associated with changes in F_b , a local F0 increase, represented by a rise in the amplitude of the tone command from the midpoint of the utterance, and a final-rise in F0 where the additional boundary tone command for the final syllable was also noted in questions. The tone of the syllable at the utterance final position was observed to have an impact on the overall F0 contour.

6. References

- Chang, C. T. (1958). Tones and intonation in Chengtu dialect. *Phonetica*, 2, 59-85.
- [2] Chang, C. Y. F. (2003). *Intonation in Cantonese*. Muchen: LINCOM Europa.
- [3] Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.
- [4] Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *Journal of Acoustical Society of America*, 80, 402-415.
- [5] Fitzsimons, M., Sheahan, N., & Staunton, H. (2001). Gender and the integration of acoustic dimensions of prosody: Implications for clinical studies. *Brain and Language*, 78, 94-108.
- [6] Fok-Chan, Y. Y. (1974). *A perceptual study of tones in Cantonese*. Hong Kong: University of Hong Kong Press.
- [7] Fry, D. B. (1968). *Manual of phonetics*. Amsterdam: North Holland Publishing Co.
- [8] Fujisaki, H., & Hirose, K. (1984). Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of Acoustical Society of Japan*, 5, 233-242.
- [9] Gu, W-T., Hirose, K., & Fujisaki, H. (2004). Analysis and synthesis of Cantonese F0 contours based on the command-response model. Paper presented at the International Symposium on Chinese spoken language processing, Hong Kong SAR.
- [10] Gu, W-T., Hirose, K., & Fujisaki, H. (2005a). Identification and synthesis of Cantonese tones based on the command-response model for F0 contour generation. Paper presented at the International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, USA.
- [11] Gu, W-T., Hirose, K., & Fujisaki, H. (2005b). Analysis of the effects of word emphasis and echo question on F0 contours of Cantonese utterances. Paper presented at the Interspeech, Lisboa, Portugal.
- [12] Henton, C. (1995). Pitch dynamism in female and male speech. Language and Communication, 15, 43-61.
- [13] Ho, A. T. (1977). Intonation variation in a Mandarin sentence for three expressions: Interrogative, exclamatory and declarative. *Phonetica*, 34, 446-457.
- [14] Ma, K-Y. J., Ciocca, V. & Whitehill, T. (2004a). The effects of intonation patterns on lexical tone production in Cantonese by acoustic analysis. Paper presented at the International Symposium on Tonal Aspects of Languages, Beijing, China.
- [15] Ma, K-Y. J., Ciocca, V. & Whitehill, T. (2004b). The effect of intonation on perception of Cantonese lexical tones. Paper presented at the International Conference on Spoken Language Processing, Jeju, Korea.
- [16] Mixdorff, H. (2004). FujiParaEditor. Retrieved November, 16, 2004, from Berlin University of Applied Sciences website:

http://www.tfberlin.de/~mixdorff/fujisaki_analysis.htm

- [17] Vance, T. J. (1976). An experimental investigation of tone and intonation in Cantonese. *Phonetica*, 33, 368-392.
- [18] Yuan, J-H. (2004). Perception of Mandarin intonation. Paper presented at the International Symposium on Chinese Spoken Language Processing, Hong Kong SAR.