ISCA Archive
http://www.isca-speech.org/archive

TAL 2006 – International Symposium
on Tonal Aspects of Languages
La Rochelle, France
April 27-29, 2006

Perception of intonation 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

This study investigated the significance of different intonation markers in intonation perception, and the effect of lexical tone on the perception of intonation. Six tonal contrasts at the final position were presented to twelve listeners in three presentation conditions: complete sentence, carrier-only and final-syllable-only. Results showed similar perceptual accuracy in both complete sentence and final-syllable-only conditions, but significantly poorer performance was observed in the carrier-only condition. This showed that intonation markings at the final position are important for correct intonation identification, when compared with perceptual cues of the carrier. In addition, differences in perceptual accuracy were noted for statements and questions embedded with different tones at the final position, suggesting that lexical tone has an effect on the identity of intonation.

1. Introduction

Intonation is a universal feature, shared by different world languages, that plays a significant role in conveying both linguistic and paralinguistic meanings in communication. Intonation is marked by variation of fundamental frequency (F0) at sentential level. Previous acoustic analyses showed that statements are associated with a falling F0 contour while questions have a rising F0 contour in most of the world languages [3, 6, 8]. Acoustic analyses also showed that the F0 pattern of a question has a higher F0 level than the same sentence produced as a statement through the entire course of the sentence [3, 6].

In a previous intonation perception experiment in Mandarin, Yuan [10] presented a set of stimuli with minimal contrast of intonation to 16 native speakers for intonation identification task. He found that statements were easier to identify than questions. He suggested that statement was a default choice in intonation perception, and listeners will only perceive the stimuli as a question when the intonation markings were clearly identified. Gussenhoven & Chen [5] investigated the significance of final-rise as an intonation marker in question perception. By asking three groups of monolingual speakers to perceive stimuli with different pitch peak height and timing, they found that questions are associated with either later peak or higher peak regardless of the language background of a speaker. They also found that the perceptual patterns differed across listeners of different language background, with stimuli showing similar F0 pattern to questions in one's own language more likely to be identified as questions. They hypothesized that both innate and nonlinguistic knowledge on the meaning of pitch variation as well as one's language background affect listener's interpretation strategies in identifying intonation. However, as question was marked by both overall F0 level increase and final rise in F0, it is possible that either F0 marking can serve as a perceptual cue to the listeners. The significance of different intonation markers in perception will be examined in this study.

The perception of intonation in tone languages like Cantonese can be potentially confusing as F0 also marks lexical tones at syllabic level. For example, the identity of question may be more ambiguous when it ends with a falling tone. Previous experiments on intonation perception in Mandarin found that utterance final tone has no impact on the identification of statements, but questions ending with falling tones were easier to identify than those ending with rising tones [10, 11]. Yuan and Shih [11] suggested that as questions and rising tone share a similar contour, the rising contour of the tone may have masked the speaker's effort in producing the rising contour of questions. However, they did not explain the better performance of questions ending with falling contour.

There are six tones in Cantonese: high-level (55), high-rising (25), mid-level (33), low-falling (21), low-rising (23) and low-level (22) [2]. It is hypothesized that tones of same contour may affect intonation differently, as the magnitudes of the F0 change differ. For example, for the two rising tones (tones 25 and 23), questions ending with tone 25 may be more accurately perceived as the syllable ends with higher peak, as suggested by Gussenhoven and Chen [5]. Furthermore, previous acoustic analyses on the effect of intonation on tone suggested basic differences between Mandarin and Cantonese. While tone contour was not modified by question intonation in Mandarin [6], previous studies showed that all Cantonese tones occurring at the final position of questions end with a rising contour regardless of their canonical forms [4, 7]. The rising F0 contour in Cantonese may help the listeners in the identification of questions.

The purpose of the present study was to investigate the perceptual cues listeners used in intonation identification, and to investigate the effect of lexical tones on the perception of intonation in Cantonese.

2. Method

2.1. Listeners

Twelve females were recruited as listeners. They were all first year undergraduates in the Division of Speech and Hearing Sciences, the University of Hong Kong. Their age ranged from 18 to 19 years old. They were considered naïve listeners as the experiment was carried out within the first two months of their enrollment, during which they received limited training in phonetics. Cantonese was the native language of all the listeners. Hearing screening was performed with all the listeners by the first author. All listeners passed the hearing screening (≤ 20 dBHL at 250, 500, 1000, 2000 and 4000 Hz).

2.2. Stimuli

Speech materials were collected from two native Cantonese speakers [one male (M1) and one female (F1)]. A carrier phrase with words contrasting in tones embedded at the final position was used (/lei$_{55}$ k $_{33}$ tsi$_{22}$ h i$_{22}$ <u>X</u>/ ‘This word is <u>X</u>’). The carrier phrase can be produced as either a question or a statement. Three sets of words, derived from the syllables /j u/, /ji/ and /si/, were embedded in the above context. Each set consisted of six words that differed only in tone. With one context, eighteen words and two intonations, there were a total of 36 different stimuli produced by each speaker.

In the current experiment, three presentation conditions were used – the “complete sentence”, “carrier-only” and “final-syllable-only”. For the complete sentence condition, the original utterance produced by the speaker was presented. As the entire utterance was presented, the various perceptual cues listeners use in intonation identification would be preserved. It was expected that stimuli presented within this presentation condition would be accurately perceived. In the other two presentation conditions, either the carrier or the final syllable was presented. The carrier and the final syllable were manually extracted from the original production using the Praat software [1]. The beginning and the end of the carrier or the syllable were selected by looking at the waveform display plot and the wide-band spectrogram. These two presentation conditions were designed to investigate perceptual cues listeners used in intonation perception. Previous acoustic analyses showed that questions are marked by both a final-rise in F0 [3, 5, 8] and an overall higher F0 level of the entire sentence [3, 5]. The relative significance of various perceptual cues in intonation perception can be evaluated by presenting either the carrier or the final-syllable only.

2.3. Procedures

The experiment was carried out in a sound-attenuated room (IAC booth), with speech materials presented to listeners through a Sennheiser HD 545 headset, connected to an Apple Macintosh G4 computer. A HyperCard programme was used to run the experiment. The stimuli were divided into six blocks according to presentation condition (complete sentence, carrier-only and final-syllable-only) and talker. Within each block, there were a total of 108 trials, as each of the 36 stimuli was repeated three times. The order of presentation of blocks was counterbalanced across listeners. For each trial, two buttons labelled “Question” and “Statement” were presented on the screen. The listeners were asked to identify the intonation of the stimulus heard by clicking on one of the buttons. Each stimulus was presented once and the listener could opt to listen to it for a second time by clicking on the “repeat sound” button. Each block took about ten minutes to finish, and the whole session took about one hour for each listener.

3. Results

Perceptual accuracy was calculated as the percentage of responses identified as the intended intonation for each listener according to presentation condition (complete sentence, carrier-only and final-syllable-only), intonation (question and statement) and embedded tone (tone 55, tone 25, tone 33, tone 21, tone 23 and tone 22). The mean perceptual accuracy was compiled across the twelve listeners as shown in Table 1. A series of Wilcoxon matched pair test was used to compare the results across intonation, presentation condition and embedded tone.

	Question			Statement		
Tone	CS	FS	CO	CS	FS	CO
55	81.02	84.26	48.15	100	98.15	70.83
25	96.76	93.06	27.31	86.57	77.78	82.87
33	100	96.30	68.98	99.07	96.76	82.09
21	84.26	86.58	36.11	100	100	82.41
23	90.74	84.26	45.83	100	99.07	85.19
22	92.13	88.43	52.78	100	100	73.15

Table 1. Perceptual accuracy according to intonation (question and statement), presentation condition [complete sentence (CS), final-syllable-only (FS) and carrier-only (CO)] and tone (tone 55, tone 25, tone 33, tone 21, tone 23 and tone 22)

3.1. Intonation

The perceptual accuracy between the two intonations was compared within each presentation condition. For tones 55, 21, 23 and 22, the perceptual accuracy was significantly higher for statements than questions in all presentation conditions ($p < 0.05$ for all). For tone 33, very similar perceptual accuracy between the two intonations was noted in both the complete sentence condition and the final-syllable-only condition. Perceptual accuracy was significantly higher for statements than questions when only the carrier was presented for tone 33 ($\underline{T} = 14$, $\underline{p} < 0.05$). For tone 25, questions were more accurately perceived than statements in both the complete sentence condition ($\underline{T} = 7.5$, $\underline{p} < 0.05$) and the final-syllable-only condition ($\underline{T} = 2$, $\underline{p} < 0.05$). However, when only the carrier was presented, the perceptual accuracy was higher for statements than questions ($\underline{T} = 0$, $\underline{p} < 0.05$).

3.2. Presentation condition

The perceptual accuracy for statements and questions was compared separately across presentation conditions. For statements, similar identification patterns were noted for all tones, except for tone 25, across the three presentation conditions. Tone 21 and tone 22 showed 100% accuracy in both complete sentence and final-syllable-only conditions, while the accuracy for tones 55, 33 and 23 was very similar in the complete sentence and final-syllable-only conditions, ranging from 96.76% to 100%. The perceptual accuracy for tones 55, 33, 21, 23 and 22 were significantly lower when only the carrier was presented than comparing to both complete sentence condition and final-syllable-only condition ($\underline{p} < 0.05$ for all). For tone 25, contrary to the other five tones, perceptual accuracy was significantly lower when only the final syllable was presented (mean = 77.78) than in the complete sentence condition (mean = 86.57) ($\underline{T} = 5$, $\underline{p} < 0.05$). The perceptual accuracy for tone 25 in the carrier-only condition (mean = 82.87) was significantly lower than for the complete sentence condition ($\underline{T} = 5$, $\underline{p} < 0.05$), but significantly higher than when only the final syllable was presented ($\underline{T} = 17$, $\underline{p} < 0.05$).

For questions, like the statements, similar perceptual accuracy was noted both when the complete sentence was presented or when only the final syllable was presented, for all tones ($\underline{p} > 0.05$), except for tone 33. For tone 33, perceptual accuracy was significantly lower for only the final syllable (mean = 96.30, SD = 5.47) than for the complete sentence condition (mean = 100) ($\underline{T} = 0$, $\underline{p} < 0.05$). Stimuli with only the carrier presented were found to be significantly less accurately perceived than those presented in either the complete sentence condition or final-syllable-only condition for all six tones ($\underline{p} < 0.05$). The perceptual accuracy in the carrier-only condition was compared with the calculated chance level. Tone 33 has a perceptual accuracy above chance level ($\underline{p} < 0.05$); while tones 25 and 21 have perceptual accuracies below chance level ($\underline{p} < 0.05$ for both).

3.3. Tone

The effect of lexical tone on the identity of intonation was investigated by comparing the perceptual accuracy across the stimuli embedded with the six tones in the complete sentence condition. For statements, tones 55, 33, 21, 23 and 22 showed similar perceptual accuracy ($\underline{p} > 0.05$ for all), while tone 25 had significantly lower accuracy than all the other five tones ($\underline{p} < 0.05$). For questions, tones 33 and 25 were the most accurately perceived, with statistically significant higher accuracy when compared with the other four tones, except for the difference between tones 25 and 22 ($\underline{T} = 11.5$, $\underline{p} > 0.05$). Besides, tone 55 had significantly lower accuracy than tones 25, 33, 23 and 22 ($\underline{p} < 0.05$ for all), but not tone 21 ($\underline{p} > 0.05$). However, when the perceptual pattern of each individual stimulus was examined, it was found that one of the tone 55 stimuli produced by M1 was perceived with very low accuracy, while all tone 55 stimuli produced by F1 were perceived with 100% accuracy.

4. Discussion

This study investigated the perception of intonation in Cantonese, and the effect of lexical tones on intonation perception. In the current experiment, statements were more accurately perceived than questions for all tones in all presentation conditions, except for tone 25 in complete sentence and final-syllable-only conditions. The results were consistent with previous findings in Mandarin that statements are easier to perceive than questions [10]. Yuan (10) suggested that statement is a default choice for listeners in intonation perception, and that listeners would only perceive otherwise when there is enough information, e.g. a final-rise in F0, marking the intonation. Therefore, listeners may have more difficulty in correctly identifying question intonations than statements, especially when the intonation markings are ambiguous.

As intonation markings are important for correct identification of intonations, one of the main aims of this study was to evaluate the significance of different perceptual cues in marking intonations. Previous acoustic analyses showed that questions are marked by two major F0 changes: increase in F0 level of the entire sentence [3, 6] and final rise in F0 contour [3, 6, 8]. In the current experiment, the significance of these two cues was evaluated by presenting the stimuli in two presentation conditions: carrier-only and final-syllable-only. Results showed that stimuli with final syllable only had similar perceptual accuracy to those presented in complete sentences. However, when only the carrier was presented, listeners showed significantly lower perceptual accuracy than in the complete sentence condition. This finding suggests that listeners were able to correctly identify the intonation of a sentence based on F0 cues provided by the final syllable, i.e. the final-rise in F0 contour. By contrast, the perceptual cues carried by the carrier, i.e., the rise in the F0 level of the utterance, was not effective. Gussenhoven & Chen [5] showed that listeners associated higher peak and later peak in an utterance with questions, regardless of their language backgrounds. They hypothesized that this is related to the non-linguistic and innate knowledge of humans, as outlined by the “Frequency Code” [9]. According to this idea, higher frequency ending utterances appear more questioning, appealing and so on; while low frequency ending utterances seem more assertive, authoritative and so on [9]. In agreement with this hypothesis, the final-rise in F0 cued listener’s correct identification of questions in the present study. For the carrier of the questions, although the F0 level of the carrier was more raised than the same carrier produced as statements, the distinction in F0 level was not sufficient for the listeners to reliably mark questions. However, the significance of the rise in F0 level as a perceptual cue cannot be ruled out. Tone 33 in the carrier-only condition of questions was perceived at above chance level, suggesting that listeners were able to perceive the intonation correctly with the perceptual cue present in some of the carriers. Further investigation in the acoustic pattern of these stimuli might reveal what the perceptual cues are that listeners used in correctly identifying the question intonation.

Performance differed for the six tones in the complete sentence condition, suggesting that tones at the final position of an utterance have an effect on the identity of intonation. The results of the current experiment suggest that statements with tone 25 at the final position were less accurately perceived than those with the other five tones. In contrast, questions ending with tone 25 were one of the most accurately perceived, while those with tone 23 had a significantly lower perceptual accuracy. Stimuli with tone 21, the only falling tone in Cantonese, showed a similar perceptual accuracy to tone 23. These findings differ from Yuan’s [10], who found that statements are not affected by tone in perception, and that questions ending with falling tones are easier to perceive than those ending with a rising tone in Mandarin. This difference is probably related to the different effects of intonation on tone between Mandarin and Cantonese. In Mandarin, Ho [6] found that although the contour of the tone at the final syllable of question was modified, the basic shape of the tone remained similar to the isolation form. Therefore, in perceiving questions with falling tone ending, as the F0 contour of falling tone in questions does not drop as low as in statements, the F0 difference in syllable ending may have marked the question intonation more prominently. By contrast, in the case when questions end with a rising tone, as suggested by Yuan and Shih [11], the coincidence of the intonation contour and tone contour masked the speaker’s effort in marking questions. However, in Cantonese, all tones at the final position of questions have a rising F0 contour regardless of their canonical forms [7]. While tone 25 has the highest F0 level towards the end of the syllable, tones 21, 23 and 22 showed overlapping rising F0 contour at the final position of questions. The high F0 ending of tone 25 probably helped the listeners in identifying the final-rise in questions. However, a rising tone is not necessarily easier to identify as shown by the significantly lower perceptual accuracy of tone 23, which showed overlapping contour with tones 21 and 22. For these three tones, the listeners might have misperceived the overlapping rising contour as utterances ending with rising tones instead of the final F0 rise in questions, resulting in significantly lower perceptual accuracy for question intonation for these stimuli. The rising contour of tone 25 may have also confused the listeners in statement perception, as the listeners associated the high peak towards the end of an utterance with questions. However, this confusion was only observed for statements with tone 25, but not tone 23. As the main difference between the two rising tones is on the extent of the F0 rise, this suggested that listeners do not perceive all sentences ending with a rising F0 as questions, but only those with a high F0 peak at the end.

5. Conclusions

This study examined the relative importance of different intonation markers in intonation perception. Listeners relied more on cues provided by the final syllable than those in the carrier in the identification of intonation. Tones at the final position of syllable have a significant effect on intonation identity, which depends on the magnitude of F0 changes carried by lexical tone. Questions ending with high-rising tone were more accurately perceived than those ending with low-rising tone, while statements ending with high-rising tone were more likely to be misperceived.

6. References

[1] Boersma, P. & Weenink, D. (2003). *Praat Program.* Retrieved February, 27, 2003, from University of Amsterdam, Institute of Phonetics Sciences website: http://www.fon.hum.uva.nl/praat/

[2] Chao, Y. R. (1947). *Cantonese primer*. New York: Greenwood Press.

[3] 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.

[4] Fok-Chan, Y. Y. (1974). *A perceptual study of tones in Cantonese*. Hong Kong: University of Hong Kong Press.

[5] Gussenhoven, C., & Chen, A-J. (2000, October). *Universal and language-specific effects in the perception of question intonation.* Paper presented at the International Conference on Spoken Language Processing, Beijing, China.

[6] Ho, A. T. (1977). Intonation variation in a Mandarin sentence for three expressions: Interrogative, exclamatory and declarative. *Phonetica, 34*, 446-457.

[7] Ma, J. K-Y, Ciocca, V. & Whitehill, T. (2004, March). *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.

[8] Ohala, J. J. (1978). Tone Rules. In V. A. Fromkin (Ed.), *Tone: An linguistic approach* (pp. 5-39). New York: Academic Press.

[9] Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols & J. J. Ohala (Eds.), *Sound symbolism* (pp. 325-347). Cambridge, UK: Cambridge University Press.

[10] Yuan, J-H. (2004, December). *Perception of Mandarin intonation.* Paper presented at the International Symposium on Chinese Spoken Language Processing, Hong Kong SAR.

[11] Yuan, J-H., & Shih, C-L. (2004, March). *Confusability of Chinese intonation.* Paper presented at the international conference of Speech Prosody, Nara, Japan.