Cross-Linguistic Perception of Intonation by Mandarin and Cantonese Listeners

Bo Robert Xu, Peggy Mok

Department of Linguistics and Modern Languages, The Chinese University of Hong Kong

xuborobert@gmail.com, peggymok@cuhk.edu.hk

Abstract

The question intonation patterns of Mandarin and Cantonese are distinctive. While Mandarin relies on the global F0 contour, Cantonese depends on a boundary tone. A cross-linguistic perception test was conducted with Mandarin and Cantonese listeners to investigate their use of pitch in identifying intonation patterns in familiar and unfamiliar languages. Forced identification tasks were used. The perception results confirmed previously established intonation patterns of the two languages. The results also revealed the universal features of pitch that listeners use to differentiate intonation types. However, language specific and tone specific factors also play a part in the perception process. Therefore, perception of intonation is an interaction of the universal biological process and language-tone specific process.

Index Terms: intonation, cross-linguistic perception, Cantonese, Mandarin

1. Introduction

The prosodic features of Mandarin and Cantonese were shown to be different in previous acoustic studies. First, the lexical tone system of Cantonese is more complicated than that of Mandarin. Second, they use distinctive intonation patterns to signal questions. This study investigates the cross-linguistic perception of intonation by listeners of these two languages.

1.1. The Lexical Tones of Mandarin and Cantonese

Figure 1 shows the four lexical tones in Mandarin, each with a distinctive contour (level, rising, dipping and falling).

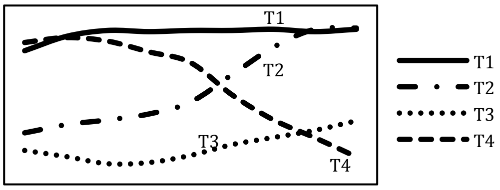

Figure 1: *Lexical tones in Mandarin (time normalized).*

The tone system in Cantonese is more complicated. Six tones contrast in both tone shape and pitch height, with a crowded tone space especially in the lower pitch range. There are three tones in the high register (T1 level, T2 rising, T3 level), and three in the low register (T4 falling, T5 rising, T6 level).

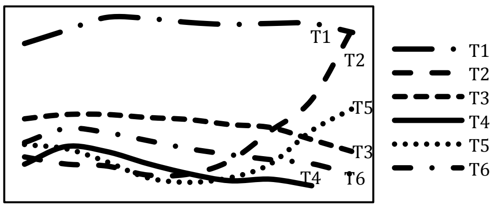

Figure 2: *Lexical tones in Cantonese (time normalized).*

The Cantonese and Mandarin tone systems share a high level tone (T1 in both systems) and a high rising tone (T2 in both systems). However, it is possible that listeners will perceive them differently in the intonational context of the other language.

1.2. The Intonation of Mandarin and Cantonese

Mandarin and Cantonese manipulate pitch differently to signal questions. It was reported that questions in Mandarin were cued by a raised global F0 contour. A boundary tone was unnecessary. The shapes of lexical tones were not affected in any position of a question ([1][2][3]). On the contrary, Cantonese questions did not rely on a global pitch raise. Instead, a boundary tone H% results in a final rising. All lexical tones except T1 in the final position of a question were affected and had a rising tail ([4][5][6]).

1.3. Perception of Intonation in Both Languages

Yuan and Shih have found several perceptual asymmetries in Mandarin intonation ([3][7][8]). First, statements were easier to identify than questions. Second, the identification of statement was not affected by the last tone while question was. Third, questions ending with T4 are the easiest to identify, whereas T2 the most difficult. Studies have also shown that the placement of focus played a role in statement-question identification ([3][9]).

Ma et al. [10] studied the perception of Cantonese intonation by comparing the perception of complete questions, final syllables and questions where the final syllables were cut off. Their results showed that even though listeners could make use of intonation cues in the global pitch contour, final rising remains the critical factor contributing to intonation perception. In their study, Cantonese speakers also displayed a perceptual bias towards statements. They also showed perceptual confusion between lexical tone and intonation, particularly when final rising coincided with the rising lexical tones (T2, T5).

1.4. Summary

Given the different prosodic features in Mandarin and Cantonese, a legitimate question to ask is whether speakers of these languages can perceive the distinction between statements and questions in the other language. A cross-linguistic perception test conducted by Gussenhoven and Chen [11] showed that three different groups of monolingual listeners (Dutch, Chinese and Hungarian) displayed similar association between question intonation and either a later or a higher F0 peak in a made-up language. This concurs with the Frequency Code proposed by Ohala ([12][13]), which claimed that pitch is innately associated with certain pragmatic meanings including questioning. However, even though both languages utilize a high pitch as a signal of questioning, Mandarin employs pitch level over a sentential scale, whereas Cantonese uses a high rising boundary tone (over canonical lexical tones) that is strictly localized in the sentence final position. The different scope and the different use of pitch (register raise vs. rising) would potentially cause differences in the production and perception of intonation. A cross-linguistic perception study will offer unique insights into this issue

Furthermore, complicated tone-intonation interactions occur in both Mandarin and Cantonese. A cross-linguistic investigation would provide valuable tools to explore the psychological reality of such interactions. So far, there is no study conducted between tone languages using controlled speech materials. This study will examine such relationship by comparing the identifications of intonations in Cantonese and Mandarin.

2. Method

2.1. Speech Materials

2.1.1. Stimuli Design

The materials in both languages were designed to compare the effects of final rising and global contour on the perception of question intonation, and the interaction between lexical tones and intonation. Accordingly, two sets of nine-syllable sentences in Mandarin and Cantonese shown in Tables 1 and 2 were included in the experiment. The final two syllables of each sentence share the same tone. With the final syllable cut off, the utterances still remained meaningful, and the ending tone remained the same (with the exception of T3 in Mandarin because of tone sandhi which is unavoidable).

Table 1. Mandarin utterances used in the experiment.

Finals	Sentences in *pinyin* with English translation
T1	*ma1 ma jin1 wan3 dun4 de shi4 ji1 (tang1).* ‘Mommy cooked chicken (soup) for tonight’
T2	*ya3 ma3 xun4 shi4 zu4i chang2 de he2 (liu2).* ‘Amazon is the longest river.’
T3	*ta1 zui4 da4 de que1 dian3 shi4 lan3 (san3).* ‘His biggest shortcoming is laziness.’
T4	*gong1 ren2 zai4 xiu1 gong1 yuan2 de lu4 (mian4).* ‘The workers are repairing the road in the park.’

Table 2. Cantonese utterances used in the experiment.

Finals	Sentences in *Jyutping* with English translation
T1	*maa1 maa1 gam1 maan1 zyu2 ge3 hai6 gai1(tong1).* ‘Mommy cooked chicken (soup) for tonight’
T2	*kam4 jat6 lou5 wong2 keoi5 dei6 hou2 zou2 (zau2).* ‘Yesterday Mr. Wong and his friends leave early.’
T3	*keoi5 kam4 maan5 saam1 dim2 sin1 heoi3 fan3 (gaau3).* ‘He went to bed at three last night.’
T4	*aa3 maa5 seon3 hai6 zeoi3 zeong4 ge3 ho4 (lau4).* ‘Amazon is the longest river.’
T5	*aa3 po4 kam4 yat6 maai5 ge3 hai6 haai5 (lau5).* ‘Granny bought crab (stick) yesterday.’
T6	*daai6 hok6 haau6 jyun4 jau5 hou2 do1 syu6 (muk6).* ‘There are many trees on campus.’

2.1.2. Recording and editing

Two native Hong Kong Cantonese (1 M, 1 F) and two native Beijing Mandarin speakers (1 M, 1 F), were recorded reading the sentences for the experiment. They were either graduate or undergraduate students at CUHK, aged between 19 and 27. The recording took place in a sound-treated room, where random sentences appeared on a computer screen. Each subject read the sentences in two forms, as a question and as a statement, with a previous instruction. To eliminate the potential discrepancy caused by the different focus patterns in the two languages, the speakers were instructed to read the sentences focus-neutrally.

After screening the naturalness of the utterances produced by the native speakers, four presentation sentence conditions (complete statements and cut-off statements, complete questions and cut-off questions) were prepared for the perception test. All cutting points were at zero crossing. The average amplitude of all the utterances was normalized using PRAAT.

2.2. Listeners and Procedures

Fifteen Cantonese and fifteen Mandarin listeners were paid to join the experiment. They were MA or undergraduate students at CUHK, between 18 and 24 years old. All the Cantonese listeners were native Hong Kong Cantonese speakers, able to speak Mandarin with varying proficiency. All the Mandarin listeners came from Mainland China, speaking Standard Mandarin in their daily life. They had been in Hong Kong for less than two months and still found Cantonese difficult to understand. All the listeners spoke English as a second language. None of them had a reported history of speech or hearing disorder.

The perception experiment was carried out in a sound-attenuated room. The materials were presented on a computer screen in random order. Each of the 80 stimuli (4 sentence conditions × 2 genders × 4 tones in Mandarin and 6 tones in Cantonese) was repeated twice, resulting in 160 stimuli in total. Participants listened to four blocks; each with 40 Cantonese and Mandarin utterance mixed up. Task instructions were given visually on screen and verbally by an experimenter. Listeners were allowed to listen to each trial repeatedly before marking down whether the utterance they heard was a statement or a question on an answer sheet.

3. Results

Identification accuracy (IA) was calculated as the percentage of correct identification of the sentence type given. The identification accuracies of both groups of listeners are shown below. In each figure, the numbers in the horizontal axis stand for the lexical tones; L stands for complete sentences; C stands for cut-off sentences.

3.1. Cantonese listeners listening to Cantonese

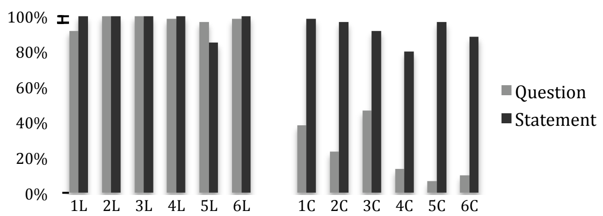

Figure 3: *IA of Cantonese utterances by Cantonese listeners.*

Figure 3 shows the identification accuracy of Cantonese intonation by Cantonese listeners. As native speakers, it is not surprising that they performed well in classifying complete utterances for both statements and questions. They occasionally failed to identify questions ending with T1 (91%) and statements ending with T5 (85%). However, when listening to sentences with the final syllable cut off, they had much more difficulty in identifying questions while the identification of statement remained relatively more successful. The questions ending with tones in the higher register (T1, T2, T3) appeared to give more cues to the listeners, as they are identified more accurately.

3.2. Mandarin listeners listening to Cantonese

Figure 4 shows the identification accuracy of Cantonese intonation by Mandarin speakers. When listening to complete utterances, Mandarin listeners could generally identify the sentence type, even though they inevitably performed worse than native Cantonese speakers. Mandarin listeners had difficulty with questions ending with T1 (68.3%) and T6 (78.3%), statements ending with T5 (66.7%) and T6 (55%). The cut-off utterances also posed problems for Mandarin listeners, as the accuracy dropped. Mandarin listeners shared the same pattern with Cantonese listeners: they did better in identifying cut-off utterances ending with tones in the higher register. Surprisingly, Mandarin listeners did better in cut-off questions ending with T2 (56.7%) and T5 (28.3%) than Cantonese listeners (23.3% and 6.7%), probably due to Mandarin listeners' misinterpretations of the rising contour.

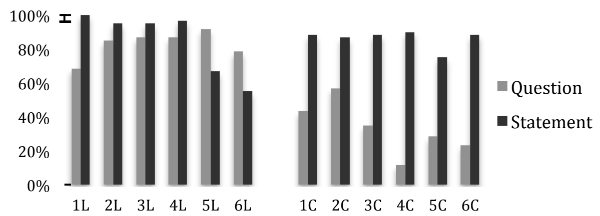

Figure 4: *IA of Cantonese utterances by Mandarin listeners.*

3.3. Mandarin listeners listening to Mandarin

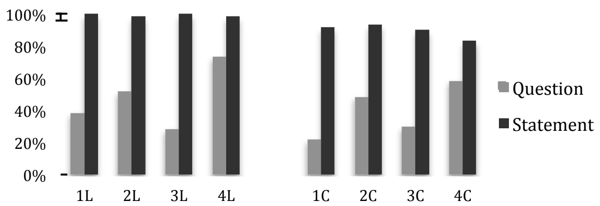

Figure 5: *IA of Mandarin utterances by Mandarin listeners.*

Figure 5 shows the identification accuracy of Mandarin intonation by Mandarin listeners. It was obvious that they did not perform as well as Cantonese native speakers on Cantonese utterances (see Figure 3), especially for questions. However, general patterns compatible with previous studies can be observed. Most noticeable is the higher accuracy of questions ending with T4 (73.3% in complete utterances and 58.3% in incomplete ones) than T2 (51.7% and 48.3% respectively). It is also interesting that when sentence ending was cut off, while the overall accuracy of statement type identification dropped, the accuracy of question identification remained similar to that of complete sentences.

3.4. Cantonese listeners listening to Mandarin

Figure 6 shows the identification accuracy of Mandarin intonation by Cantonese listeners. Probably due to the subjects' higher familiarity to Mandarin than Mandarin listeners to Cantonese, they did relatively well in this task, even better than native Mandarin listeners in identifying statements when utterance endings were cut off. However, Cantonese listeners also had some difficulties. First of all, they could not recognize questions ending with T3 in either condition (0% in complete utterances and .05% in incomplete ones). They showed reverse patterns in identifying questions ending with T2 and T4 (T2 better than T4) in comparison to Mandarin listeners (see Figure 5).

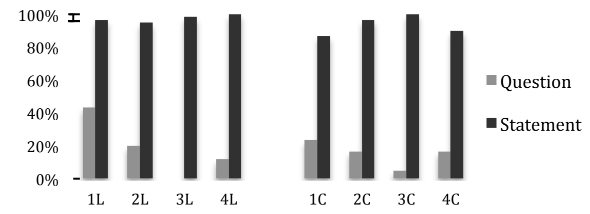

Figure 6: *IA of Mandarin utterances by Cantonese listeners.*

4. Discussion

The results confirm the intonation patterns of the two languages investigated in previous studies. Without the final syllable, where a boundary tone is located, the accuracy of identification of questions in Cantonese decreased dramatically. This result is compatible with acoustic studies ([4][6]) claiming that a boundary tone located at the very end of the final tone is the critical cue for Cantonese questions. This could also explain the relatively low intonation accuracy of T1 complete questions, as T1 is less likely to become rising at the end of questions due to a ceiling effect since T1 is at the top of a speaker's pitch range already ([4][6]). In contrast, the accuracy of identification was not significantly affected by losing the final syllable in Mandarin, showing that the signal of Mandarin question is not a local event, but the global contour. However, after removing the final syllable of Cantonese questions, both Cantonese and Mandarin listeners could still make some correct judgments, indicating that some intonation information can still be found in the remaining part of the utterances. In Xu and Mok ([6]), T3 was found to be the most consistent tone to exhibit a global raise of pitch level in questions in Cantonese. This can partly explain why Cantonese listeners scored the best in identifying Cantonese T3 cut-off questions. Furthermore, the poor performance of Mandarin listeners in spotting questions in either complete or cut-off utterances in their own language is also on a par with previous studies on the parallel encoding of focus and question intonation ([9]).

The results further the insights into question-statement bias previously observed in single languages. For both groups of listeners in both languages, the question identification accuracy was constantly lower than statement, especially in cut-off utterances. On the other hand, statement accuracy was consistently higher than 80%. This indicates that statement was the preferred choice when strong acoustic cues for questions were not present. The fact that both groups of listeners performed so similarly in the two languages suggests that this bias (statement being an unmarked sentence type) may be universal.

Gandour's pioneer study [14] on Cantonese tonal perception pointed out that listeners can employ different dimensions of pitch to identify linguistic contrast, e.g. contour, direction and height. Our study has provided additional evidence to his suggestion. On hearing the Cantonese cut-off questions, listeners of Cantonese and Mandarin both made most mistakes with the tones in the low register (T4, T5, T6), while performed better with tones in the high register (T1, T2, T3). This suggests that listeners related the high register with question intonation regardless of the pitch shape of the final tones they hear. The height dimension of pitch was prominent enough to overshadow the contour dimension. That was why listeners of both languages performed better with a level tone (Cantonese T1 high level and T3 mid level) than a low rising tone (T5). Another piece of evidence comes from the perception of T2 (a high rising tone in both languages). While both groups of listeners performed poorly in T2 questions in their own language, they performed well in the language that they were not familiar with, even better than the native listeners. This is probably due to the association in their mind between a high rising contour and question intonation, when they were not familiar with the prosodic features of the language they heard, and in the case of many Mandarin speakers, when semantics was absent.

The cross-linguistic results also lend some support to Ohala's proposal of Frequency Code ([12][13]), which stated that smaller vocal cords and faster vibrations rates were used for expression of power relations: a high and high-ending pitch corresponds to appealing and questioning, a low and low-ending pitch corresponds to authority and assertiveness. As discussed above, a high-rising and high register pitch contours were more often identified with questions, when some critical information like boundary tone was cut off. The identification of questions with low dipping tone (T3) in Mandarin was rather poor by Cantonese listeners. This further supports a general tendency for languages to exploit high and high rising pitch to convey questions.

While the use of universal code can be supported in the data, some language-specific patterns could also be observed. The perceptual asymmetry between T2 and T4 in Mandarin has been reported in previous studies. This study confirms that the falling tone (T4) makes questions in Mandarin easiest to identify, while the rising tone (T2) very difficult. Yuan ([3]) argued that this is because intonation perception was sensitive to tonal identity. As shown in this study, when such tonal identity was absent, even when the contours were largely the same (Cantonese T2 and Mandarin T2 are both high rising tones), Mandarin listeners had the best identification accuracy in T2 Cantonese cut-off questions. In turn, Cantonese listeners also identified T2 questions in Mandarin best, while they did not do so well in their own language. The new cross-linguistic asymmetry observed in this study shows that language-specific tonal awareness is more important than the interpretation of general intonation tendency. When hearing their native language, listeners tend to identify the lexical tones before intonation patterns.

Finally, some facts cannot be accounted for by prosodic theories, such as the low IA of T6 complete statement in Cantonese by Mandarin listeners. The following two reasons may explain why: (a) the interference of syllable structure and duration. The sentence ended with an entering tone (syllable ending with unreleased stop) due to difficulties in stimuli design. Mandarin listeners might treat the unusually short duration and the unreleased quality differently, and thus identified T6 statements as questions. Or (b) the interruption of lexical factors, as some Mandarin speakers thought that the last syllable sounded like a question particle in their own language (as reported by some Mandarin listeners). Nonetheless, the interplay between prosody and other aspects of language warrants further study in this interesting area.

5. Conclusion

Our cross-linguistic investigation has provided some new understanding of the processing of tone-intonation perception and the dimensions of pitch employment in tone languages. In addition to confirming previous studies on the perceptual patterns of intonation in Cantonese and Mandarin, this study has also extended these observations from a universal perspective. Different dimensions of pitch used in Mandarin and Cantonese can be observed in the data, in which high-rising and high-level pitch contours were associated with question intonation in less familiar languages.

However, the perception of universal pitch patterns of question intonation can also be influenced by language-specific and tone-specific factors. The perception of intonation, as a result, is likely an interaction of the two processes.

6. References

[1] Connell, B.A., Hogan J.T., and Rozsypal, A.J. 1983. Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *Journal of Phonetics* 11, 337-351.

[2] Yuan, J., Shih C. and Kochanki, G.P. 2002. Comparison of declarative and interrogative intonation in Chinese. *Proceeding of the Speech Prosody*, Aix-en-Provence, France, 711-714.

[3] Yuan, J. 2004. *Intonation in Mandarin Chinese: Acoustics, Perception and Computational Modeling.* Ph.D. Thesis. Cornell University

[4] Ma, J.K., Ciocca V. and Whitehill T.L. 2006a. Effect of intonation on Cantonese lexical tones. *Journal of Acoustic Society of America* 120(6), 3978-3987.

[5] Ma, J.K., Ciocca V. and Whitehill T.L. 2006b. Quantitative analysis of intonation patterns in statements and questions in Cantonese. *Proceedings of the Third Intonational Conference on Speech Prosody*, Dresden, Germany, 277-280.

[6] Xu, B. R. and Mok, P. 2011. Final rising and global raising in Cantonese intonation. *Proceeding of the International Congress of Phonetic Sciences.* Hong Kong. 2173-2176.

[7] Yuan, J. and Shih, C., Confusability of Chinese intonation, *Proceedings of the Speech Prosody 2004*, pp. 131-134.

[8] Yuan, J. Perception of Mandarin intonation. *Proceedings of International Symposium of Chinese Spoken Language Processing*. Hong Kong. 45-48.

[9] Liu, F. and Xu, Y. 2005. Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica* 2005:62. 70-87.

[10] Ma, J.K., Ciocca V. and Whitehill T.L. 2010. The perception of intonation questions and statements in Cantonese. *Journal of Acoustic Society of America* 129(2). 1012-1023.

[11] Gussenhoven, C and Chen Chen, A. 2000. Universal and language-specific effects in the perception of question intonation. *Proceedings of the International Conference on Spoken Language Processing*. Beijing, China. 91-94.

[12] Ohala, J.J. 1983. Cross-language use of pitch: an ethological view. *Phonetica* 40. 1-18.

[13] Gussenhoven, C. 2002. Intonation and interpretation: phonetics and phonology. *Proceedings of the Speech Prosody 2002. Aox-en-Provence.* 47-57.

[14] Gandour, J. 1981. Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics*, 9. 20-36.