Intonation Perception of Low-Pass Filtered Speech in Mandarin and Cantonese

Xu Bo Robert, Peggy Mok

Department of Linguistics and Modern Languages, The Chinese University of Hong Kong
xuborobert@gmail.com, peggymok@cuhk.edu.hk

Abstract

The intonation patterns of Mandarin and Cantonese are very different. While Mandarin uses a raised global F0 contour to signal questions, Cantonese relies on a boundary tone strictly located at the end of a question. Previous individual-language studies and cross-linguistic studies with normal speech stimuli have shown that intonation interpretation was influenced by both universal perceptual factors and language-tone specific factors. In this study, Mandarin and Cantonese listeners participated in a forced identification task to judge the sentence type of the low-pass filtered Mandarin or Cantonese utterances they heard. The results confirm the different intonation patterns between Cantonese and Mandarin. They also show that when lexical-semantic information was absent and when identities of lexical tones were covered up, listeners relied solely on the frequency code to interpret the intonation. Finally, the results also suggest that F0 contour alone was not sufficient for intonation perception. Lexical-semantic context, segmental information and duration-rhythmic profile also contribute to intonation production and perception.

Index Terms: intonation, perception, low-pass filter, Mandarin, Cantonese

1. Introduction

Mandarin and Cantonese are two closely related tone languages spoken in Chinese communities. However, their prosodic features have been shown to be different in two ways. First, Cantonese employs a much more complicated lexical tone system than Mandarin. Second, they use distinctive intonation patterns to signal questions and focus. This research followed our previous study to investigate the cross-linguistic perception of intonation by Mandarin and Cantonese listeners in low-pass filtered speech. The aim was to find out how listeners interpreted the intonation of both languages (their native language and a less familiar language) when lexical and segmental information was absent.

1.1. The lexical tones in Mandarin and Cantonese

Four lexical tones in Mandarin are contrasted by only tone shape. T1 is a high level tone, T2 is a high rising tone, T3 is a dipping tone in the low register, and T4 is a high falling tone.
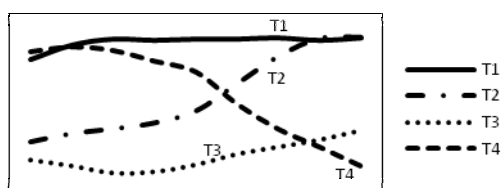

Figure 1: *Mandarin Lexical tones (time normalized).*

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).
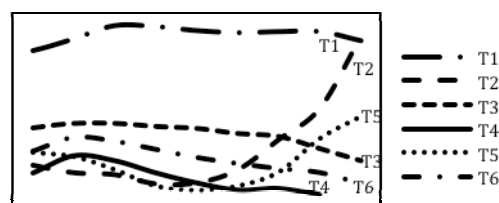

Figure 2: *Cantonese Lexical tones (time normalized).*

1.2. Intonation of Mandarin and Cantonese

1.2.1. Production

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% at the end of a question results in a final rising. As shown in Figure 3 (taken from [14]), all lexical tones except T1 in the final position of a question were affected and had a statistically significant rising tail, whereas the global raise was neither significant nor consistent [4][5][6].
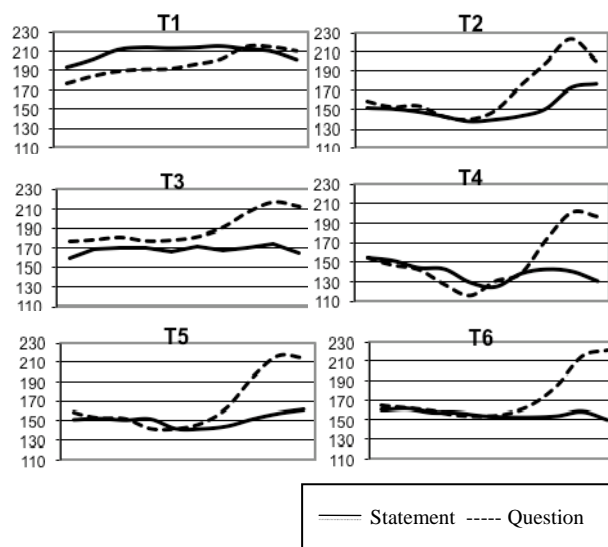

Figure 3: *Sentence final tones of questions and Statements in Cantonese (taken from [14]).*

1.2.2. Perception

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. Yuan [3] argued that this was because intonation perception was sensitive to tone identity. Studies have shown that the placement of focus also 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).

1.3. Cross-linguistic intonation perception

While intonation patterns remain diverse in languages, cross-linguistic intonation perception was rarely studied. 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 was innately associated with certain pragmatic meanings including questioning.

A previous study concerning Mandarin and Cantonese by the current authors [14] tested the cross-linguistic perception of intonation in normal speech. The results were in accordance with the different intonation patterns of Cantonese and Mandarin found in previous production tests. It also confirmed perception patterns in previous research that focused on one of the languages. However, the study revealed that listeners interpreted similar F0 contours differently in distinct intonation contexts. For example, when listening to their native language, both Cantonese and Mandarin listeners showed difficulties in identifying questions ending with a high rising lexical tone. On the contrary, they identified such questions better in the other language, as predicted by the Frequency Code. This suggested that the interpretation of intonation was an interaction between universal perceptual processes and language-tone specific processes.

1.4. Summary

Even though both languages utilize a high pitch in their intonation contour as a signal of questioning, the high pitch is realized differently. Mandarin employs pitch level over a sentential scale, whereas Cantonese uses a boundary tone that is strictly localized in the sentence final position, which results in a high rising over canonical lexical tones. The different scope and the different use of pitch (register raise vs. rising shape) have been shown to cause differences in both the production and perception of intonation. Our previous cross-linguistic study with Mandarin and Cantonese [14] used normal speech stimuli to show the interaction between universal Frequency Code and language-tone specific perception. This study applied low-pass filter to the speech materials in order to find out whether listeners would rely on the universal Frequency Code entirely to perceive intonation when they could not recognize the semantic context.

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, in which case the T3 dipping tone would be realized as a high rising contour).

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. Low-pass filters were applied to every sound file with 100 Hz bandsmoothing. The cutoff frequency varied according to the individual pitch range, determined by the highest pitch each

speaker produced in their questions with T2 (high rising tone for both languages). The cutoff frequency was 200 Hz for the Cantonese male speaker, 320 Hz for the Cantonese female speaker, 150 Hz for the Mandarin male speaker, and 270 Hz for the Mandarin female speaker. Informal tests showed that native speakers of both languages could not understand the content of the sentences at all. They reported only hearing low frequency humming. Last, 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 participate in the perception 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, and could 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. 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 on the horizontal axis stand for the lexical tones; L stands for complete sentences; C stands for cut-off sentences.

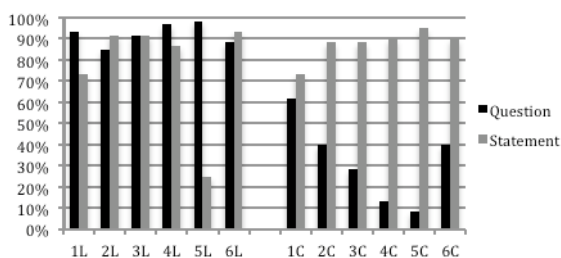## 3.1. Cantonese listeners listening to Cantonese

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

Figure 3 shows the identification accuracy of Cantonese intonation by Cantonese listeners. For the complete utterances, listeners could generally identify the sentence type, with the exception of statements with T1 (73.3%) and T5 (25.0%), due to their high register (T1) and rising shape (T5) respectively. For utterances with the ending cut off, the IA was significantly lower for all questions, while IA for statements stayed similar with the complete statements. Listeners showed difficulties in recognizing both questions and statements for T1 utterances

without the final tones (61.7% and 73.3% respectively). They also had troubles identifying questions with the other tones (T2 40.0%, T3 28.3%, T6 40.0%), among which T4 and T5 questions were the worst (13.3% and 8.3% respectively).

3.2. Mandarin listeners listening to Cantonese

Figure 4 shows the identification accuracy of Cantonese intonation by Mandarin listeners. The result has a resemblance to that of Cantonese listeners. For complete utterances, participants did well in all tones, especially in questions. Like Cantonese speakers, they also had problems with statements with T1 (60.0%) and T5 (26.7%). For utterances with the final tones cut off, they had similar IA in statements as they did in complete statements. But they did not do as well in the questions. No IA for questions with endings cut off surpassed 65%, with T2 being the highest (63.3%) and T4 the lowest (38.3%). However, in every tone, Mandarin listeners performed better in identifying questions without the final tones than did Cantonese listeners. It was also interesting that in complete utterances with T1 to T5, Mandarin listeners performed better in identifying questions than statements, while the IA for complete questions and statements with T6 was the same. This shows that the final rising in Cantonese was probably a strong interrogative cue even for non-native listeners, and was strongly associated with questions.
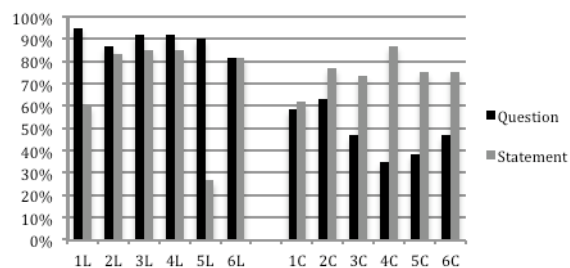

Figure 4: *IA of Cantonese utterances by Mandarin listeners.*

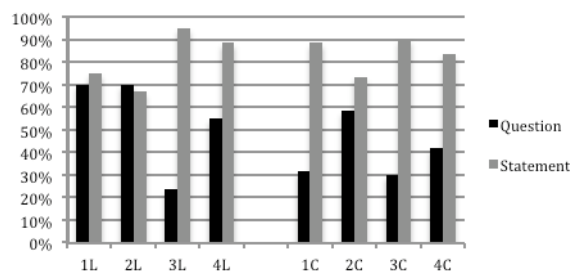## 3.3. Mandarin listeners listening to Mandarin

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

Figure 5 shows the identification accuracy of Mandarin intonation by Mandarin listeners. They had more troubles recognizing questions than statements. For T2 to T4, getting rid of the final tone did not seem to have a great impact on the identification of sentence type. For T1, when listening to complete utterances, Mandarin listeners had similar accuracy for statements and questions; however, when they listened to utterances without the final tones, they could identify statements much better (75.0% up to 88.3%) than questions

(70.0% down to 31.7%). In contrary to how they performed with non-filtered speech ([3][14]), they identified questions ending with T2, a rising tone, better than questions ending with T4, a falling tone, in both complete and cut-off utterances. On the other hand, in both conditions, they did better in T4 than T2 for statements. They struggled the most with questions ending with T3, in both complete utterances (23.3%) and utterances without the final tones (30.0%). However, they were most successful at identifying statements ending with T3 among all tones and conditions.

3.4. Cantonese listeners listening to Mandarin

Figure 6 shows the identification accuracy of Mandarin intonation by Cantonese listeners. For T1 and T4, Cantonese listeners scored higher in identifying questions in complete utterances (73.3% and 48.3% respectively) than in cut-off utterances (33.3% and 25.0% respectively). For T2, they identified the same amount of questions successfully in the two conditions (41.7%). And for T3, which underwent tone sandhi and became a rising tone in sentences with the final tone cut off, they had better IA in cut-off utterances (35.0%) than in complete utterances (10.0%). In general, however, Cantonese listeners did not perform well in Mandarin questions as they identified all questions ending with all tones and under both conditions below chance level (50.0%), with the only exception of T1 in complete utterances (73.3%). In complete utterances, Cantonese listeners did slightly better for questions ending with T4 (48.3%) than those ending with T2 (41.7%). In contrast, in utterances with the final tones cut off, they did better with questions ending with T2 (41.7%) than T4 (25.0%).
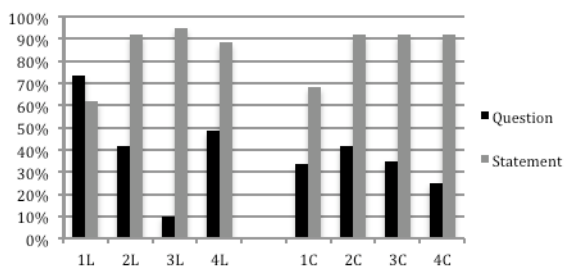

Figure 6: *IA of Mandarin utterances by Cantonese listeners.*

4.   Discussion

The perceptual results of this experiment reflected the acoustic differences between Mandarin and Cantonese intonation found in previous production studies. In Cantonese, where questions were signaled by a boundary tone locating at the very end of the last syllable, cutting off the final tones of questions significantly affects the perceptual accuracy of questions in all tones. The identification accuracies of Cantonese listeners were brought down to below chance level (50.0%) in all tones except T1 (73.3%), probably owing to the fact that T1 was the only tone that does not have a significant final rising in questions due to its being at the top of the speakers' normal pitch range. Mandarin listeners also did poorly with cut-off questions in Cantonese, with the accuracy of T3 to T6 below chance level. They did relatively better in T1 (58.3%) and T2 (63.3%), probably because those were the two lexical tones in the high register and thus were more easily associated with questions. However, previous studies showed that the general contour could provide some, even though limited, resources

for question identification in Cantonese [10][14]. Given their own language utilizes global contour to signal questions, Mandarin listeners might be able to capture such trivial cues in Cantonese with better sensitivity. This could probably explain why Mandarin listeners performed better in cut-off Cantonese questions than did Cantonese listeners.

On the other hand, cutting off the final tone of Mandarin utterances did not affect listeners as much as abridging Cantonese utterances did. The IA of listeners did not drop very dramatically after cutting off the final tone in Mandarin. There were, however, two exceptions. For both Cantonese and Mandarin listeners, T3 questions were identified more accurately in cut-off utterances than in complete utterances. This was because, as shown in Figure 7, in cut-off utterances, the ending tone (second to last tone of the complete utterance) underwent tone sandhi and became a rising tone, which could easily be associated with questions. Questions ending with T1 were identified more successfully in complete utterances, for both groups of listeners. This was because, as shown in Figure 8, the F0 contours of the final two T1 (high level tones) were realized differently. While the last T1 (the last tone in the complete question) was realized as a high level tone, the second to last T1 (the last tone in the cut-off question) was realized with a falling tail. As a result, the falling tail at the end of the sentence was more likely to be associated with statements even though the register of the whole utterance was raised because of question signaling.
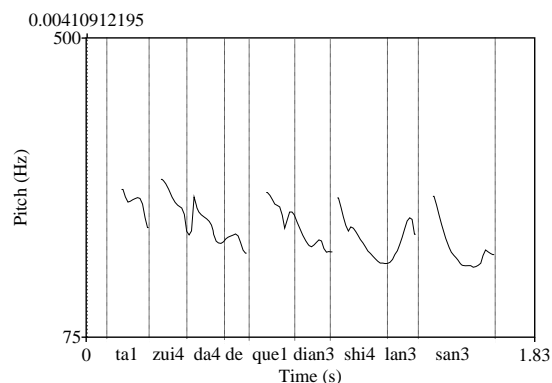

Figure 7: *The F0 contour of the complete T3 question by the Mandarin female speaker used in the experiment.*
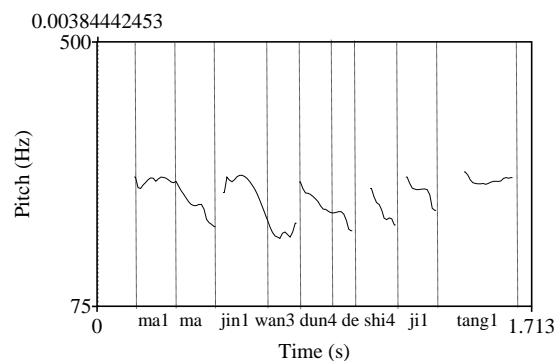

Figure 8: *The F0 contour of the complete T1 question by the Mandarin female speaker used in the experiment.*

Besides, the results also supported the statement-question bias found in previous studies [3][14], as statements were generally identified more accurately than questions, especially in all utterances with final tones cut off. This showed that listeners tended to label a sentence as a statement when strong cues for question intonation were absent.

In comparison to our previous study using normal speech stimuli [14], the general identification accuracy is lower in the current study. This suggests that listeners found it more difficult to identify sentence type with filtered speech materials. Since the low-pass filter eliminated most segmental information, we can conclude that lexical information and sentential meaning contribute to intonation interpretation. Furthermore, the low-pass filter made individual syllables, rhythmic patterns and speech rate in the utterances less clear to the listeners, which might also have negative effects on intonation perception. Therefore, this result invites further studies on the interactions in intonation perception between F0 contour and other linguistic components such as lexical information, semantic context and supplementary prosodic features.

Since the low-pass filter made individual tones and language difficult to recognize, the language-tone specific effect found in previous studies [3][14] was greatly weakened. The effect of the Frequency Code was expected to dominate listeners' decision regardless of the native language they spoke and the language they heard. The results seemed to support this prediction. First, the identification accuracy patterns between Cantonese and Mandarin speakers were in general very similar. This indicated that listeners were less influenced by their top-down phonological knowledge of intonation, but instead relied more on the F0 contour and the universal code. Second, the perception patterns of Mandarin T1 and T3 question discussed above showed that a rising contour (even though starting from a low pitch) caused by the T3 sandhi would bring up the identification accuracy, while a falling shape (even though in a high register) realized in a non-final T1 would lower the identification accuracy. Third, the T2-T4 bias in Mandarin intonation perception by Mandarin listeners, as shown in previous studies [3][14], was not found in this study. In contrast, Mandarin listeners had a higher accuracy with T2 questions than T4 questions in both complete and cut-off utterances. This showed that their interpretation was affected by the Frequency Code when low-pass filter was applied. Further, this implied that the language-tone specific factors absent in this study due to the filter played an important part in intonation perception of tone languages. Fourth, Mandarin questions ending with T3 (dipping tone in a low register) in complete utterances were constantly misidentified by both Mandarin and Cantonese listeners. Likewise, the identification accuracies of cut-off questions ending with tones in the mid to low register in Cantonese (T3 to T6) were lowest for both groups of listeners. These facts lent further support to the Frequency Code, as a low tone was clearly not associated with question signaling.

Another issue in the results requires further discussion. For both Mandarin and Cantonese listeners, the T5 statements in the complete utterances were poorly identified. Some may argue that this was due to the rising contour of T5 in Cantonese. However, given that T5 is a low rising tone in Cantonese and T2 is a high rising tone, one would predict that the T2 statements should result in an even lower accuracy than T5 statements. Nonetheless, T2 statements were well perceived by both groups of listeners. Therefore, the rising contour alone could not account for the low accuracy of T5 complete statements. A close look at the acoustic pattern of the F0 contour revealed that the complete T5 statement

(Figure 9) ended with a steeper rise than the complete T2 statement (Figure 10).
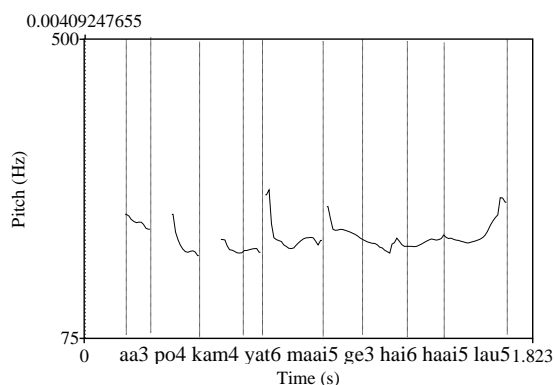

Figure 9: *The F0 contour of the complete T5 statement by the Cantonese female speaker used in the experiment.*
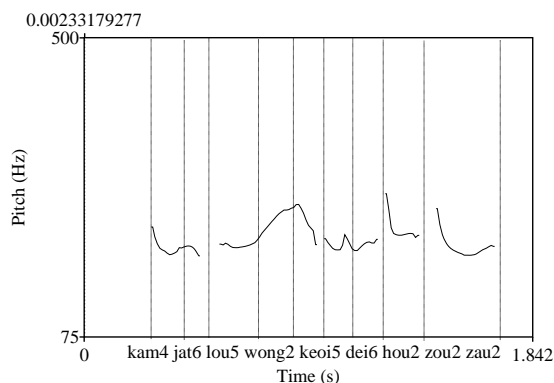

Figure 10: *The F0 contour of the complete T2 statement by the Cantonese female speaker used in the experiment.*

The reasons for this pattern are as follows. First, Cantonese is currently undergoing tone merging. Some young speakers do not clearly distinguish T2 and T5 in their production. This phenomenon affects lexical items differently, resulting in unstable production of some T5 words. After reexamining the Cantonese speakers in this study, it was found that the speakers produced the last T5 word of the complete utterance (*lau5*) with a higher than usual rise, approaching that of T2. But they produced the second to last T5 word (*haai5*) with a canonical T5. Therefore, the ending of the complete T5 utterance sounded like a high-rising T2 rather than a canonical T5. Second, even though the final tones of the T2 and T5 utterances were both high rising, the F0 contours were further modified by declination and tonal coarticulation. In the case of the T2 statement (Figure 10), as the pitch range narrowed down towards the end because of declination, the tail of the final T2 had a limited rising. In addition, a carry-over effect brought by the previous high rising tone causes the onset of the final tone to be even higher than the rising tail. As a result, the final tone of the T2 statement did not have a dramatic rising and was not necessarily associated with questions by the both groups of listeners. On the other hand, in the T5 statement, the final tone was preceded by a low rising tone with only a minimal rise, which renders the uncannoical rise of the final T5 very salient. This prominent rising at the end of a sentence was associated with questions by the listeners, and

resulted in a low IA for the T5 statements. To sum up, the final tones of T2 and T5 statements were affected by tone merging. However, they appeared differently because of declination and the influence of the previous tones, resulting in a conspicuous rise (T5) and an inconspicuous one (T2). The results further support the Frequency Code: statements ending with a conspicuous rise were misidentified as questions, and statements ending with a mild rise were identified as statements.

5. Conclusion

This study examined cross-linguistic perception of intonation of low-pass filtered speech by Mandarin and Cantonese listeners. When lexical and semantic information was absent, as well as when lexical tone identity was difficult to recognize, listeners relied solely on the information provided by the sentential F0 contour. The results showed that their interpretation of intonation was dominated by the frequency code. The high register and a rising contour were associated with question signaling, whereas the low register and a falling contour were the signal for statements. The results also showed that F0 contours resulting from language-specific reasons such as tone sandhi and tonal coarticulation contribute to intonation perception, especially when top-down phonological knowledge could not be applied. This study has confirmed the differences between Cantonese and Mandarin intonation with filtered speech. The fact that both Mandarin and Cantonese listeners performed below chance level in identifying Cantonese questions without the final tone clearly indicate that the boundary tone in Cantonese is critical for question signaling. While in Mandarin, a boundary did not seem to be perceptually necessary. Finally, the results also imply that F0 was not the only source of intonation information in normal speech. Further investigation into the interactions among F0 contour, lexical-semantic context, segmental information, and the prosodic profile is needed in order to have a more comprehensive understanding of intonation in different languages.

6. References

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

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

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

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

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

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

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

[8] Yuan, J. Perception of Mandarin intonation. Proceedings of International Symposium of Chinese Spoken Language Processing, 45-48, 2004.

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

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

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

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

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

[14] Xu, B. R. and Mok. P. K. Cross-linguistic perception of intonation by mandarin and Cantonese listeners. Proceedings of the Speech Prosody, 2012.