Declarative and Interrogative Mandarin Intonation by Native Speakers and Cantonese L2 Learners

Wentao Gu and Lei Liu

Nanjing Normal University, China

wtgu@njnu.edu.cn, 1990liulei@163.com

Abstract

This study compared sentence intonation of L1 Mandarin by native speakers and L2 Mandarin by Cantonese learners, with both acoustic analysis and perceptual experiment. Three types of sentences (i.e., statement, intonation question, and particle question) ending with different tones and in different lengths were investigated. The perceptual experiment showed that declarative intonation in L2 speech was better identified than in L1 speech, which could be explained by the more prominent F_0 declination in L2 speech. In contrast, interrogative intonation in L2 speech had a lower rate of identification than in L1 speech, and the differences in rate varied with the sentence-final tone. Acoustic analysis showed that global F_0 raising for questions was weaker in L2 speech than in L1 speech, especially in longer sentences, while sentence-final F_0 raising was relatively well maintained in L2 speech. Perceptual and acoustic studies showed consistent results on L2 intonation errors, which could be explained by the limited language abilities and language transfer effects.

Index Terms: declarative, interrogative, intonation, Mandarin, Cantonese, L2 speech

1. Introduction

The naturalness of speech depends highly on its prosody including tone and intonation. The acoustic manifestations of tone and intonation, as well as the manner how they interact with each other depend highly on the language. A number of previous studies, e.g., [1], have shown that many L2 prosodic errors are attributed to language transfer effects resulting from these cross-linguistic differences.

Mandarin and HK Cantonese (henceforth Cantonese) are both Chinese tone languages. In terms of Chinese traditional phonology, a syllable in Chinese languages is divided into an initial consonant and a final rhyme carrying a tone which is cued mainly by fundamental frequency (i.e., F_0). Despite this common property, Mandarin and Cantonese contrast sharply in prosodic phonology for both tone and intonation.

As shown in Table 1, Mandarin has four lexical tones (T3 is a low tone, except on-focus or at the sentence-final position where it is a 'dipping' tone), and a neutral tone functioning as an unstressed syllable, in which F_0 does not have an intrinsic pattern but varies largely with the preceding tone. In contrast, Cantonese has six lexical tones, and no neutral tone.

In both languages, the statement and the unmarked yes-no question (i.e., intonation question) are associated with falling and rising sentence intonation, respectively. The pattern of F_0 raising in intonation question, however, is language-dependent. In Cantonese questions, F_0 is raised mainly at the final syllable [2-4], which can be described by a boundary tone in the AM theory [5]. Despite some debates, it is generally agreed that F_0 in Mandarin questions is raised in a longer domain, even starts at a higher level than in statements [6, 7], and the amplitude of F_0 raising increases with time, peaking at the final syllable [8].

The interaction between tone and intonation is a critical issue in tone languages. The most conspicuous interaction lies in the sentence-final syllable of a question with a rising intonation. In Mandarin intonation questions, the relative tone pattern in the final syllable is not modified, but F_0 of the entire final syllable is raised – the later the higher [9]. In Cantonese intonation questions, F_0 in the final syllable in any tone has a rising shape [2, 3], which in the framework of Fujisaki model can be modeled by replacing the tone command in the later part of the final syllable by a particular positive command [4].

The cross-linguistic differences in acoustic manifestations for tone and intonation are also reflected in the perceptual characteristics. In both Mandarin and Cantonese, statements are generally better identified than questions [10-12]. For Mandarin, some studies, e.g., [10], showed that boundary tone played limited roles in perceiving questions, which instead were cued by a global higher pitch register than statements. Some other studies, however, reported that perception of questions also depended on the sentence-final tone; e.g., easier to identify when the final tone was T4 rather than T2 [11-13], or sometimes most difficult to identify when the final tone is T3 [12, 13]. For Cantonese, previous studies showed that boundary tone played crucial roles in perception [12, 14] – listeners tended to associate high pitch register in the final syllable with question intonation regardless of its pitch contour.

Although the above differences in intonation between Mandarin and Cantonese have been noticed, there have been few studies on intonation errors in L2 Mandarin speech by Cantonese learners. A controlled acoustic experiment has recently revealed the L2 F_0 errors in statements and intonation questions [15]. In the present study, particle questions will also be considered, and sentence length will be included as another factor. In addition, elicited speech from more natural dialogues than [15] will be employed to better investigate declarative and interrogative intonations in communicative speech.

Table 1: *Tone systems of Mandarin and Cantonese.*

Mandarin			Cantonese		
type	feature	value	type	feature	value
T1	high	55	TC1	high level	55
T2	rising	35	TC2	high rising	25
T3	low/dipping	21(4)	TC3	mid level	33
T4	falling	51	TC4	low falling	21
			TC5	low rising	23
T0	neutral	–	TC6	low level	22

2. Speech data

For the purpose of a controlled comparison, we designed three sets of sentences, for which the Chinese text, the pinyin transcription, and the direct English translation are shown below:

(a) 今天吃[煎包 / 鲜桃 / 尖枣 / 酸酪][。/？/吗？]

"Jin1 tian1 chi1 [jian1 bao1/ xian1 tao2/ jian1 zao3/ suan1 lao4] ./?/ma0?"

Today (we) eat [fried buns/ fresh peach/ tsim jujube/ yoghurt] ./?

(b) 今天吃苏州煎包[。/？/吗？]

"Jin1 tian1 chi1 su1 zhou1 jian1 bao1 ./?/ma0?"

Today (we) eat Suzhou fried buns ./?

(c) 今天张哥吃苏州煎包[。/？/吗？]

"Jin1 tian1 zhang1 ge1 chi1 su1 zhou1 jian1 bao1 ./?/ma0?"

Today Brother Zhang eats Suzhou fried buns ./?

These sentences end with a period, or a question mark with or without a preceding question particle /ma/ (in T0), representing a statement, a marked or unmarked yes/no question (i.e., a particle question or an intonation question), respectively. The particle and intonation questions share the same meaning. For the convenience of description, the syllable immediately before the particle /ma/ in particle question is also termed 'sentence-final' syllable.

Sentence set (a) has a fixed carrier frame, in which four disyllabic target words (names of foods) shown in brackets are embedded. The first syllable in the target word shares similar rhymes /ian/ or /uan/, while the second syllable shares a fixed rhyme /ao/. All syllables in sentence set (a) have a high tone T1, except the sentence-final syllable which varies in four tones. This design aims to examine intonation without the interaction of lexical tones except the boundary tone. Sentence sets (b) and (c) use a fixed target word /jian1 bao1/, but they consist of 7 and 9 syllables (without counting the particle /ma/), respectively, all of which are in T1. This design aims to examine the effect of sentence length.

All these sentences are meaningful texts, varying systematically in three factors:

- Sentence type (3 levels): statement, unmarked yes-no question (i.e., intonation question), and marked yes-no question (i.e., particle question).
- Lexical tones in the sentence-final syllable (4 levels): T1~T4; neutral tone was not considered here.
- Length of sentence (3 levels): 5, 7, or 9 syllables.

It should be noted that the 7- and 9-syllabic sentences end only with T1. Thus, there are altogether (4+2)×3 = 18 sentences.

Two groups of informants participated in the experiment. They were native in Mandarin and HK Cantonese, respectively. Each group consisted of ten informants (5M+5F) at similar ages – the average ages for the L1 and L2 groups are 25 and 19, respectively. The L1 informants were all graduate students with a high proficiency level of Mandarin, while the L2 informants were HK learners of Mandarin at the medium level – they had studied Mandarin at university for one or two years.

To examine intonation variations in natural speech, the present study adopted elicited speech in a role-play. For each target sentence, either statement or question, we designed a dialogue between two parties. Each dialogue consisted of 3 to 6 turns, and the target sentence constituted a single turn by itself. The prompt texts were also provided to elucidate the scenario and the relationship between the two parties.

Speech recording was done in a sound-proof room after the informants got familiar with the materials and felt certain with the pronunciation of all the texts. The dialogues between two parties were conducted in a conversational style, and were monitored by the experimenter. Once there was any apparent mistake or disfluency, the informants would be asked to repeat recording the dialogue until success.

3. Perceptual experiment

To explore the perceptual characteristics of L1 and L2 intonation, we conducted a perceptual experiment, in which 10 native speakers of Mandarin (5M+5F) were recruited as listening subjects. They were all graduate students around the age of 24, without any reported hearing impairments. There was no overlap between the recording informants and the listening subjects.

All 18×20 = 360 target utterances extracted from the recorded dialogues were used for perceptual judgment of statement vs. question. For particle questions, the particle /ma/ was cut away from the utterances to ensure that all perceptual judgments were based only on prosody. Although the truncated utterances sounded incomplete, a subjective prediction of statement vs. question could still be made.

The E-Prime software was used for stimulus presentation and response collection. The method of constant stimuli was adopted as the test paradigm. All 360 stimuli were combined randomly into 12 sound files, each composed of 30 stimuli with an inter-stimuli interval of 5 seconds. These sounds were presented to the subjects through headphones in a sound-proof room. Within each 5s inter-stimuli interval, the subjects were requested to judge the sentence type by choosing from three options: 'statement', 'question', and 'unsure'. If the subjects failed to respond within the time interval, the system would take 'unsure' as the default answer. The assignment of response keys was counter-balanced across listening subjects. Before the experiment, a training session was repeated until the listening subjects got used to the procedure and could give the answers confidently.

Figure 1 shows the rates of perceptual identification of sentence type for all 5-syllabic sentences ending with four different tones. The identification rates for 7- and 9-syllabic sentences ending with T1 gave a similar pattern as those for their 5-syllabic counterparts, so they are not plotted here.

Largely in line with the results in previous studies [10-12], the rates of identification for statements are in most cases higher than for intonation questions, for both the L1 and L2 groups. The only exception showing an obviously lower rate of identification for statement than for its question counterpart exists in the L1 statement ending with T2, suggesting that even L1 speech sounds a bit confusing when sentence intonation and the final tone conflict in the direction of pitch movements.

Also, in most cases, the rates of identification are much lower in truncated particle questions. This is predictable because interrogation in particle questions is conveyed mainly by the final particle – it is hard to make judgments when the particle is cut away. The only exception lies in the L1 truncated particle question ending with T4, which gives a comparable rate of identification as its statement counterpart. This suggests that the regressive effect of the T0 final particle on the preceding tone is perceptible only in the case of T4.

A comparison between the L1 and L2 speech shows that for statements L2 speech is better identified than L1 speech (except when the final tone is T4), which will be explained later after acoustic analysis. On the contrary, for both types of questions L1 speech is consistently better identified than L2 speech, indicating that the L2 group has not acquired the prosodic coding strategy for Mandarin questions successfully.

For both L1 and L2 speech, the perception of declarative and interrogative intonations is highly influenced by the sentence-final tone. Especially, statements give the lowest rate of identification when ending with T2 (rising), whereas intonation questions give the lowest rate when ending with T4 (falling). This again suggests that perceptual judgment becomes difficult when sentence intonation and the sentence-final tone conflict in the direction of pitch movements.

More importantly, the differences in perceptual accuracy between the L1 and L2 speech are also dependent on the final tone. The relative decrease in the rate of identification for the L2 speech is most prominent (>25%) in the truncated particle question ending with T4, and the intonation question ending with T2 or T4. Therefore, among complete utterances, the perceptually most prominent L2 intonation errors exist in the intonation question ending with rising or falling tones T2/T4.

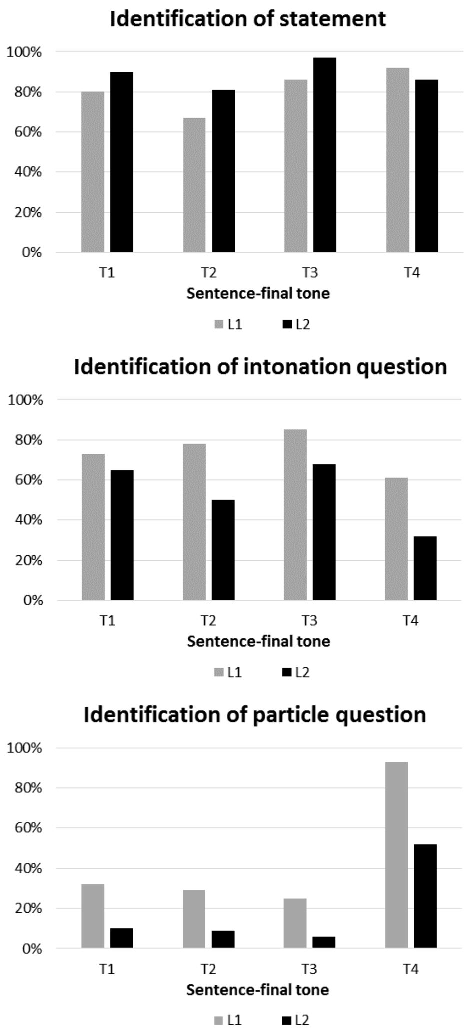

Figure 1: *Rate of perceptual identification for the 5-syllabic sentences ending with four different tones.*

4. Acoustic analysis

All 360 target utterances were segmented into syllable initials and rhymes manually. The raw F_0 values were extracted at 10ms intervals using an autocorrelation analysis in Praat. After manual correction of gross errors, F_0 values were smoothed, and then interpolated within syllable rhymes where there were breakpoints. Ignoring durational differences, the syllable rhyme based time-normalized F_0 contours were obtained by extracting F_0 values at 10 equally-spaced points in the rhyme of each syllable, and then were averaged among a certain set of informants and utterances in the scale of semitone.

Figures 2–3 show the average time-normalized F_0 contours measured in semitone. The solid and dashed lines indicate F_0 contours for the L1 and L2 groups, respectively. Three colors

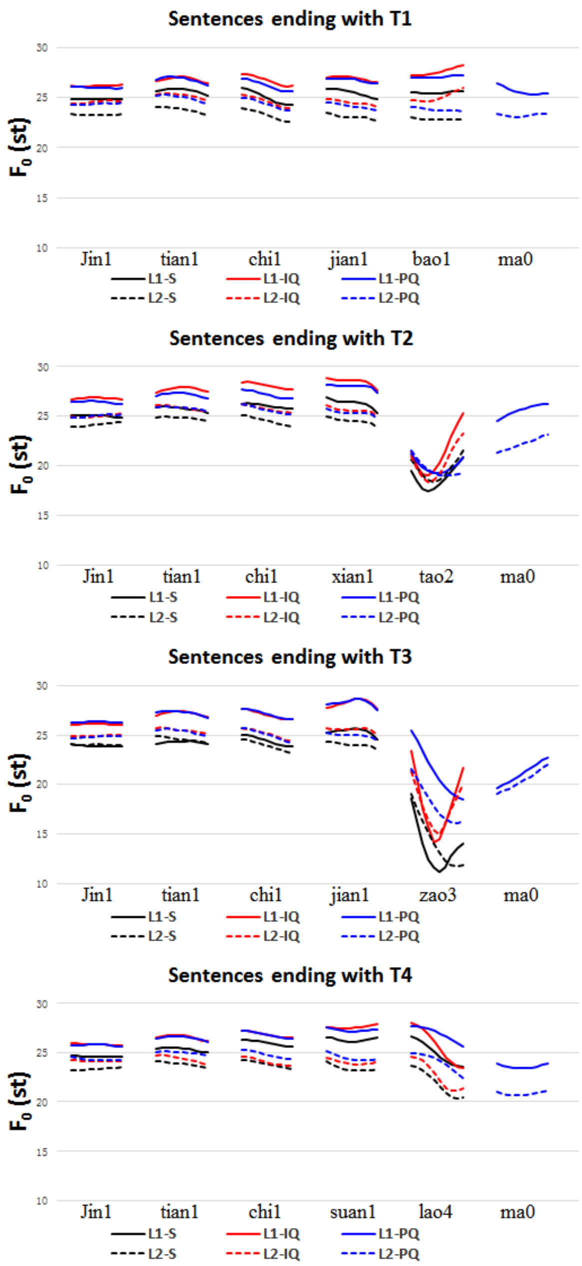

Figure 2: *Average F_0 contours for the 5-syllabic sentences ending with four different tones.*

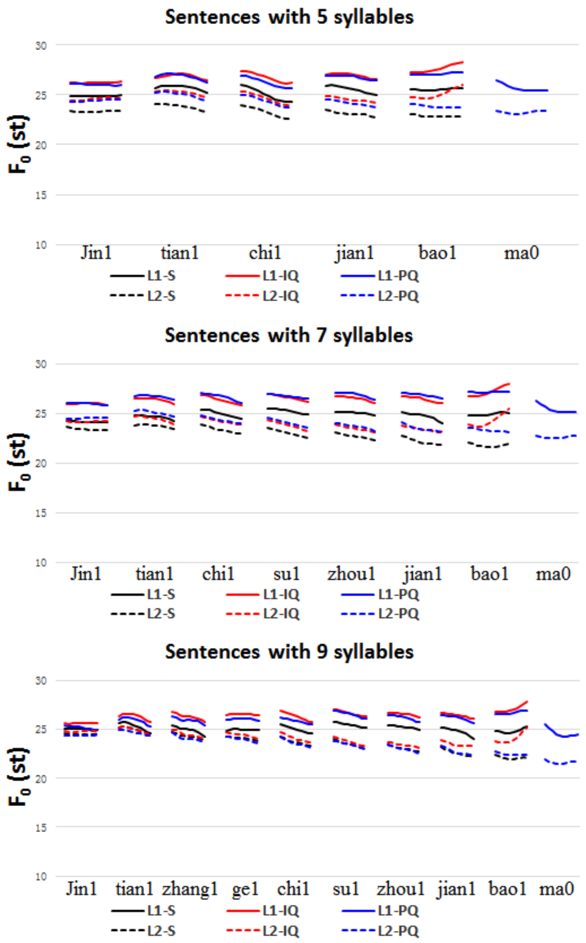

Figure 3: *Average F_0 contours for the T1-ending sentences with three different lengths.*

are used to differentiate statement, intonation question, and particle question. Based on Figs. 2–3, the L2 intonation errors can be revealed by a comparison between L1 and L2 speech.

First of all, the L2 group exhibits a conspicuously lower F_0 register than the L1 group. This coincides with our subjective impression that Cantonese speech tends to have a lower pitch than Mandarin, which may also be inferred from previous acoustic studies [16, 17]. The results here indicate that the language-dependent pitch characteristic can be transferred to L2. Also, for statements, a comparison of F_0's in all T1 syllables in the sentence shows a relatively declining tendency in the L2 speech rather than in the L1 speech.

These patterns can be exhibited more clearly in Table 2, which shows the average F_0 decrements from the L1 speech to the L2 speech in each syllable. F_0's are relatively lowered in all syllables in the L2 speech. Moreover, the amplitude of F_0 lowering in all high-tone T1 syllables increases monotonously with the time course (excluding the final syllables in other tones as shown in shadow), suggesting a relative pattern of F_0 declination in L2 declarative speech.

This finding can be explained by the fact that native speakers use F_0 to encode a variety of information on the syntactic, semantic, and pragmatic layers, and thus they may reset F_0 frequently to adjust the prosodic structure. Therefore, the F_0 contours of native speakers do not simply follow the F_0 declination rule on the articulatory layer. In contrast, the L2 group cannot code high-level information efficiently through F_0 due to their limited language abilities, and thus their F_0 contours are more dominated by the articulatory rules. This difference in F_0 declination can also explain the higher rate of identification of statements for the L2 than for the L1 group.

More conspicuous L2 errors in sentence intonation can be observed in questions. It is known that questions, especially intonation questions, generally have higher F_0 than statements, though the details of F_0 raising vary with languages. For 5-syllabic sentences ending with four different tones, Tables 3 and 4 show the average F_0 increments in each syllable, from statements to intonation questions, and to particle questions, respectively. For T1-ending sentences with three different lengths, Tables 5–6 show the same measurements. The levels of statistical significance are also indicated in Tables 3–6.

For the L1 group, in both intonation and particle questions, higher F_0 values relative to statements are shown over the entire utterance, while the amplitude of F_0 raising reaches the maximum in the sentence-final syllable in most cases, except for T4-ending intonation question and T2-ending particle question. Especially, the exceptional result for T4-ending intonation question is due to the constraint of the falling tone –

Table 2. *Syllabic mean F_0 decrements (in st) from the L1 speech to the L2 speech for the statement sentences.*

Length	Final tone	1	2	3	4	5	6	7	8	9
5-syl	T1	1.3	1.4	2.3	2.5	2.7				
	T2	0.6	1.3	1.7	2.3	0.8				
	T3	1.2	1.3	1.6	2.5	2.5				
	T4	1.4	1.9	2.6	3.4	0.7				
7-syl	T1	0.7	0.9	1.6	2.2	2.4	2.5	3.1		
9-syl	T1	0.5	0.4	0.6	0.9	1.3	2.0	2.2	2.2	2.7

Table 3. *Syllabic mean F_0 increments (in st) from statement to intonation question for the 5-syllabic sentences ending with four different tones.*

Group	Final tone	jin1	tian1	chi1	••an1	•ao
L1	T1	$1.3^{\dagger}$	1.1	1.7^{*}	$1.4^{\dagger}$	2.1^{*}
	T2	1.8^{*}	2.0^{*}	2.0^{*}	2.2^{*}	2.5^{*}
	T3	$2.2^{\dagger}$	2.9^{*}	$2.7^{\dagger}$	$2.9^{\dagger}$	4.5^{*}
	T4	$1.2^{\dagger}$	$1.2^{\ddagger}$	0.9	$1.3^{\dagger}$	$0.8^{\dagger}$
L2	T1	$1.2^{\dagger}$	$1.4^{\dagger}$	$1.4^{\dagger}$	$1.4^{\dagger}$	2.3^{*}
	T2	0.9	1.1	$1.2^{\dagger}$	$1.1^{\dagger}$	0.7
	T3	0.9	0.9	1.2	1.5	$3.5^{\dagger}$
	T4	0.8	0.5	0.3	0.6	$0.8^{\dagger}$

* $p < 0.01$; † $0.01 < p < 0.05$; ‡ $0.05 < p < 0.08$.

Table 4. *Syllabic mean F_0 increments (in st) from statement to particle question for the 5-syllabic sentences ending with four different tones.*

Group	Final tone	jin1	tian1	chi1	••an1	•ao
L1	T1	$1.2^{\dagger}$	$1.1^{\dagger}$	$1.2^{\dagger}$	$1.2^{\dagger}$	1.6^{*}
	T2	$1.4^{\dagger}$	$1.4^{\dagger}$	$1.1^{\dagger}$	1.7^{*}	$1.2^{\ddagger}$
	T3	$2.4^{\dagger}$	$3.0^{\dagger}$	$2.7^{\dagger}$	$2.9^{\dagger}$	7.7^{*}
	T4	$1.1^{\dagger}$	$1.1^{\dagger}$	$0.8^{\dagger}$	0.9^{*}	1.9^{*}
L2	T1	$1.1^{\dagger}$	$1.2^{\dagger}$	$1.1^{\ddagger}$	$1.1^{\dagger}$	$1.0^{\dagger}$
	T2	0.9	1.0	$1.1^{\dagger}$	0.8	0.1
	T3	0.8	0.8	1.1	0.9	$3.8^{\dagger}$
	T4	$1.0^{\dagger}$	$1.2^{\dagger}$	$1.0^{\dagger}$	1.1^{*}	2.1^{*}

* $p < 0.01$; † $0.01 < p < 0.05$; ‡ $0.05 < p < 0.08$.

Table 5. *Syllabic mean F_0 increments (in st) from statement to intonation question for the T1-ending sentences with three different lengths.*

Group	Length	1	2	3	4	5	6	7	8	9
L1	5-syl	1.3†	1.1	1.7*	1.4†	2.1*				
	7-syl	1.8*	1.8*	1.4*	1.4*	1.5*	1.7*	2.3*		
	9-syl	0.6	1.0	1.4†	1.6*	1.3*	1.2*	1.3*	1.6*	2.3*
L2	5-syl	1.2†	1.4†	1.4†	1.4†	2.3*				
	7-syl	0.8	0.7	0.9‡	0.7	0.8†	1.2*	2.5*		
	9-syl	0.3	0.0	0.2	0.3	0.4	0.3	0.4	0.9‡	2.0*

* $p < 0.01$; † $0.01 < p < 0.05$; ‡ $0.05 < p < 0.08$.

Table 6. *Syllabic F_0 increments (in st) from statement to particle question for the T1-ending sentences with three different lengths.*

Group	Length	1	2	3	4	5	6	7	8	9
L1	5-syl	1.2†	1.1‡	1.2‡	1.2†	1.6*				
	7-syl	1.8*	2.0*	1.7*	1.5*	1.8*	2.1*	2.2*		
	9-syl	0.1	0.6	1.0‡	1.1†	0.8†	1.1†	0.9*	1.4*	1.8*
L2	5-syl	1.1†	1.2†	1.1†	1.1†	1.0†				
	7-syl	1.1†	1.3*	1.0†	1.0*	1.0*	1.3*	1.5*		
	9-syl	0.2	0.3	0.2	0.1	0.1	0.0	0.0	0.1	0.4

* $p < 0.01$; † $0.01 < p < 0.05$; ‡ $0.05 < p < 0.08$.

this conflict between tone and intonation has also led to the lowest rate of perceptual identification as shown in Fig. 1(b).

For the L2 group, the global F_0 raising in questions is on the whole weaker than the L1 group. Either the F_0 differences are statistically less significant, or the amplitudes of F_0 raising are smaller. This kind of L2 errors is more conspicuous in longer sentences, for the apparent reason that longer sentences require longer-domain F_0 controls which pose more difficulty on L2 learners. The relatively high F_0 raising in the final syllable, however, is more consistent with the L1 group.

These L2 errors in interrogative intonation can partly be explained by language transfer effects. In Cantonese, interrogatives are implemented with a highly localized instead of a global F_0 raising. Especially, Cantonese is much richer than Mandarin in sentence-final particles, which contribute substantially to Cantonese intonation. Cantonese learners tend to maintain this strategy in their L2 Mandarin. For interrogatives, they rely more on the final F_0 raising, and do not always produce globally higher F_0 contours.

5. Conclusion

This study compared sentence intonation of Mandarin by native speakers and Cantonese learners, from both perceptual experiment and acoustic analysis. Three types of sentences (i.e., statement, intonation question, and particle question) with four ending tones and in three lengths were investigated.

Perceptual experiment showed that declarative intonation in L2 speech was on the whole better recognized than in L1 speech, while interrogative intonation in L2 speech was consistently worse recognized than in L1 speech. Also, the differences in perceptual accuracy between L1 and L2 speech were highly dependent on the sentence-final tone. In particular, the perceptually most prominent L2 intonation errors existed in intonation question ending with T2 or T4.

Acoustic analysis was also conducted. For statements, the L2 group showed more distinct F_0 declination than the L1 group. For the two types of questions, the global F_0 raising found in the L1 speech becomes weaker in the L2 speech, especially in longer sentences; in contrast, the F_0 raising in the final syllable is fairly well maintained in the L2 speech. The L2 intonation errors observed in statements and intonation questions here are largely consistent with those reported in [15], though the results here showed more variations because the speech data we examined here are more communicative.

The relation between perceptual and acoustic attributes has been clearly observed. As shown in Fig. 2, the F_0 raising from statement to intonation question is the smallest when the final tone is T4, and correspondingly, the rate of perceptual identification for question is the lowest in the same situation.

In summary, the observed Cantonese L2 learners' errors in declarative and interrogative intonation of Mandarin can be explained by their limited Mandarin abilities and language transfer effects.

The findings in this study also have obvious pedagogical implications. If the L2 learners are aware of the common error patterns in their prosodic manifestations and the resulting perceptual confusion, they might be able to overcome these errors purposely. For example, most speakers, not only L2 learners but also native speakers including language teachers, have no knowledge about the detailed prosodic manifestations for interrogatives. If the Cantonese L2 learners of Mandarin are instructed about the differences between the two languages in the strategy of pitch raising for interrogatives, it is conjectured that they will be able to improve their interrogative intonation simply by starting the utterance at a higher pitch level. The learning effects will be even better with the aid of a visual display of the F_0 contours of their speech. A systematic investigation into L2 prosodic errors will be very helpful for the L2 education, because the knowledge of the frequently occurring error patterns can be employed directly to guide the L2 education.

6. Acknowledgements

This work is supported jointly by the National Social Science Fund of China (10CYY009 and 13BYY009), the Major Programs for the National Social Science Fund of China (13&ZD189), and the key project funded by the Jiangsu Higher Institutions' Key Research Base for Philosophy and Social Sciences (2010JDXM024).

7. References

[1] J. Trouvain and U. Gut, (eds). *Non-Native Prosody: Phonetic Description and Teaching Practice*. Berlin: Mouton de Gruyter, 2007.

[2] J. K. Ma, V. Ciocca, and T. L. Whitehill, "Effect of intonation on Cantonese lexical tones," *JASA* 120: 3978–3987, 2006.

[3] B. R. Xu and P. Mok, "Final rising and global raising in Cantonese intonation," *Proc. 17th ICPhS*, Hong Kong, pp. 2173–2176, 2011.

[4] W. Gu, K. Hirose, and H. Fujisaki, "Modeling the effects of emphasis and question on fundamental frequency contours of Cantonese utterances," *IEEE Trans. Audio, Speech & Lang. Proc*. 14: 1155–1170, 2006.

[5] D. R. Ladd, *Intonational Phonology*. Cambridge: Cambridge University Press, 1996.

[6] X. Shen, *The Prosody of Mandarin Chinese*. Berkeley: University of California Press, 1990.

[7] J. Yuan, C. Shih, and G. P. Kochanski, "Comparison of declarative and interrogative intonation in Chinese," *Proc.*

Speech Prosody 2002. Aix-en-Provence, France, pp. 711–714, 2002.

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

[9] M. Lin, *Hanyu Yudiao Shiyan Yanjiu*. Beijing: China Social Sciences Press, 2012.

[10] P. Chen and A. Jiang, "Representation of Mandarin intonations: boundary tone revisited," *Proc. 23rd North American Conference on Chinese Linguistics*, vol. 1, pp. 97–109, 2011.

[11] J. Yuan and C. Shih, "Confusability of Chinese Intonation," *Proc. 2nd Speech Prosody*, Nara, Japan, pp. 131–134, 2004.

[12] B. R. Xu and P. Mok, "Cross-linguistic perception of intonation by Mandarin and Cantonese listeners," *Proc. 6th Speech Prosody*, Shanghai, China, pp. 99–102, 2012.

[13] C. Yang and K. M. Marjorie. "The perception of Mandarin Chinese tones and intonation," *Journal of the Chinese Language Teachers Association* 45 (1): 7–36, 2010.

[14] J. K. Ma, V. Ciocca, and T. L. Whitehill, "The perception of intonation questions and statements in Cantonese," *JASA* 12 (2): 1012–1023, 2011.

[15] W. Gu, "Tone, intonation, and emphatic stress in L2 Mandarin speech by English and Cantonese learners," *Proc. 18th ICPhS*, Glasgow, UK, 2015.

[16] M. L. Ng, G. Hsueh, and C. S. Leung, "Voice pitch characteristics of Cantonese and English produced by Cantonese-English bilingual children," *International Journal of Speech-Language Pathology* 12 (3): 230–236, 2010.

[17] P. Keating and G. Guo, "Comparison of speaking fundamental frequency in English and Mandarin," *JASA* 132 (2): 1050–1060, 2012.