Tonal Aspects of Languages 2016
24-27 May 2016, Buffalo, New York

The Role of the Final Tone in Signaling Statements and Questions in Mandarin

Una Y. Chow, Stephen J. Winters

School of Languages, Linguistics, Literatures and Cultures, University of Calgary, Canada
uchow@ucalgary.ca, swinters@ucalgary.ca

Abstract

This study investigated the role of the F0 height of the final tone in signaling the difference between Mandarin statements and echo questions. We recorded native speakers producing pairs of statements and echo questions ending in the four Mandarin tones (high, rising, low, and falling). The first two syllables of the utterances comprised pairs of high and low tones, or rising and falling tones. From these pairs of tones, we extracted the speaker's F0 range at the start of the utterance. With the speaker's initial F0 range serving as a baseline, we compared the difference between statements and echo questions in two ways: 1) using the overall mean of the F0 height of the final tone of each sentence type, and 2) using the mean difference in F0 height of the final tone between the sentences in each statement-question pair. The results of ANOVAs suggest that the F0 height of the final tone is a potential cue for the identification of statements versus echo questions in Mandarin. However, since significant differences in F0 were found between only some of the final tonal pairs, native speakers must also use other prosodic cues to distinguish statements from echo questions in Mandarin.

Index Terms: F0 height, tone, intonation, echo question, statement, Mandarin

1. Introduction

Mandarin is a tone language that uses four contrastive tones to distinguish lexical meanings [1]. Its tone inventory comprises T1 (a high tone, e.g., 医 *yi1* 'doctor'), T2 (a rising tone, e.g., 姨 *yi2* 'aunt'), T3 (a low tone, e.g., 椅 *yi3* 'chair') and T4 (a falling tone, e.g., 忆 *yi4* 'memory'). Described using the five-scale pitch levels [2], T1, T2, T3 and T4 have the tonal shape of 55, 35, 21(4) and 51, respectively. According to [3], native Mandarin listeners perceive fundamental frequency (F0) primarily at the lexical level, and lexical tones can reduce the listener's sensitivity to the F0 cues at the intonational level. In a perception experiment [4], sixteen native listeners of Mandarin were asked to identify 256 pairs of statements and declarative questions produced by eight native speakers of Mandarin. The sentences were presented to the listeners one at a time in randomized order. The results showed that the tone of the final syllable affected the identification of questions but not of statements. The study found that questions ending in T4 were the easiest to identify, although the researchers could not explain why. In contrast, questions ending in T2 were the hardest to identify. Motivated by these findings, we investigated lexical tone as a potential influencing factor on the variation of the pitch height of the phrase-final intonation of statements and echo questions in Mandarin.

2. Background

Previous researchers have proposed two conflicting theories of intonation systems used to signal declarative questions in Mandarin. The first theory claims that questions have an overall higher pitch contour than statements [5]. The second theory claims that the F0 difference between statement and question intonations increases towards the end of the sentences [6]. For a recent perception study, we recorded 16 native Mandarin speakers producing pairs of statements and echo questions (e.g., *Wu2 Er4 shi4 yi1 ge4 hen3 nu3 li4 de nong2 fu1. Wu2 Er4 shi4 yi1 ge4 hen3 nu3 li4 de nong2 fu1?* 'Wu Er is a very hardworking farmer'). Our observations of the recorded data found variation in F0 height of the final tone between paired statements and questions. Figures 1 and 2 illustrate this difference with a statement-question pair ending in T1 and T2, respectively.

In Figures 1 and 2, a "baseline" is shown (in red) in each pitch track. A baseline is a reference F0 value anchored at either the maximum F0 (maxF0) or minimum F0 (minF0) of the "base" (i.e., sentence-initial disyllabic name). Each base has pitch levels of 1 and 5 in its tones (e.g., *Wu2 Er4*: T2 [**35**] and T4 [5**1**]). Due to potential coarticulation of adjacent tones, the maxF0 or minF0 of a syllable might not reflect the target F0 of its tone [7, 8, 9]. Therefore, the F0 baselines were assigned as follows. If the final tone of the sentence was T1 or T4, then the baseline was the maxF0 of the base. Otherwise, if the final tone was T2 or T3, then the baseline was the minF0 of the base.

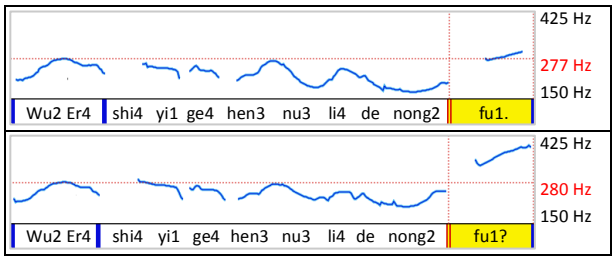

Figure 1: *'Wu Er is a very hardworking farmer', produced by a female speaker.*

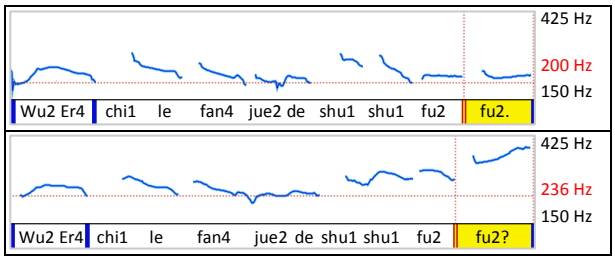

Figure 2: *'Wu Er feels comfortable after eating dinner', produced by a female speaker.*

In Figure 1, a baseline is aligned with the maxF0 of the base *Wu2 Er4* for the pair of sentences ending in T1, *fu1*. In Figure 2, a baseline is aligned with the minF0 of the base, *Wu2 Er4*, for the pair of sentences ending in T2, *fu2*. In both pairs of sentences, the F0 values of the questions are well above the baselines, while the F0 values of the statements are near the baselines.

http://dx.doi.org/10.21437/TAL.2016-36

We also observed that the amount of F0 variation in the final syllable differed among sentences that end in a different tone. For example, Figure 3 shows two statements: the first one ends in T4, and the second one ends in T2. The maxF0 of T4 is realized below the baseline, whereas the minF0 of T2 is realized on the baseline.

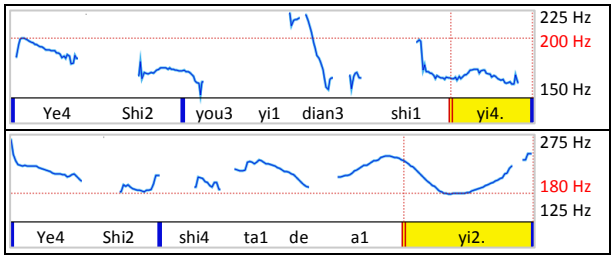

Figure 3: *'Ye Shi is slightly forgetful' (top) and 'Ye Shi is his aunt' (bottom), produced by a female speaker.*

Our research questions were the following: 1) Is the F0 height of the final syllable a cue for the identification of statements and echo questions in Mandarin? 2) Do different tones affect the F0 height of the final syllable differently in statements and in echo questions? Based on our observations above, the prediction was that the final F0 height is a potential cue for identifying statements and echo questions, and that the strength of the cue depends on the final tone.

3. Methods

3.1. Participants

This study examined a subset of the Mandarin production data which we collected for a larger research project that aimed to investigate the effects of variation on the perception of sentence intonation. In that project, eight male and eight female native speakers of Mandarin from the University of Calgary participated in a reading task. They were born and raised in China and were fluent in speaking Mandarin. They reported no history of speech or hearing disorders. The current study analyzed readings from six speakers who were randomly selected by gender: three male (aged 19, 28 and 34) and three female (aged 18, 25 and 32).

3.2. Stimuli

The stimuli for the reading task comprised five blocks of four dialogues: blocks A, B, C, D and E with target sentences that were 5, 7, 9, 11 and 13 syllables long, respectively. Every sentence in a block began with the same two-syllable name. These names included sequences of T1+T3, T4+T2, T3+T1, T2+T4 and T1+T1 tones for block A, B, C, D and E, respectively (e.g., 汪五 *Wang1 Wu3* 'Wang Wu').

Each dialogue included a target pair of Mandarin sentences, consisting of a statement and an echo question that were lexically and syntactically identical (e.g., 汪五是老师。汪五是老师? *Wang1 Wu3 shi4 lao3 shi1. Wang1 Wu3 shi4 lao3 shi1?* 'Wang Wu is a teacher. Wang Wu is a teacher?'). A filler question preceded the pair (e.g., 汪五是谁? *Wang1 Wu3 shi4 shei2?* 'Who is Wang Wu?'), thus introducing a statement reading for the first target utterance. In addition, a filler affirmative statement followed the pair (e.g., 是,汪五是老师。*Shi4, Wang1 Wu3 shi4 lao3 shi1.* 'Yes, Wang Wu is a teacher.'), thus providing a response to the second target utterance, an echo question seeking confirmation. Within each block, the target pairs ended with the same consonant and vowel sound (i.e., *shi*, *yi*, *ma*, *fu* or *fen*) but with a different Mandarin tone for each of the four dialogues. The purpose of this sentence structure was to enable us to investigate the effect of tone on the intonation of the sentence-final syllable.

This study analyzed target sentences produced with block A to D stimuli only because these stimuli began with names that had both the highest pitch level of 5 and the lowest pitch level of 1. These pitch extremes established the speaker's pitch range for the sentence and served as a base for measuring the relative F0 height of the sentence-final syllable. Altogether, this study analyzed data produced from 16 (i.e., 4 blocks x 4 dialogues) unique pairs of statements and echo questions.

3.3. Procedure

The participants were recorded individually in a sound-attenuated booth in the Phonetics Lab at the University of Calgary, using high quality recording equipment. The sentences were presented in Chinese characters in Microsoft PowerPoint on an iMac computer. Each dialogue appeared on a single slide. The speakers were instructed to read the dialogue aloud as if they were speaking in a normal conversation. They were encouraged to speak in their natural voice and at their normal speed. If they misread a sentence, they would reread the entire dialogue. The speakers read the stimuli three times, with a short break between readings. The readings were recorded at a sampling rate of 48 kHz in a 16-bit mono channel and were saved to .wav files. Only the second reading was used for the data analysis in this study.

3.4. Acoustic Analysis

To obtain the F0 values of the target sentences, the boundaries of the base and the final syllable of the recorded sentences were marked in a textgrid in Praat [10]. Then a Praat script extracted the minF0 and maxF0 values from the sound files.

If a reading error was found in a target sentence, both sentences in the pair were excluded from the analysis. In addition, sentences for which Praat failed to analyze the pitch values of the labeled intervals—due to creakiness or devoicing of the vowel—were also excluded. The final data included in the analysis totaled 73 pairs of statements and echo questions.

To test for any significant F0 difference between statements and echo questions for each of the four final tones, we performed two tests: an individual-sentence test and a paired-sentence test. For these tests, we defined a measure, a "final height" (FH), as the F0 height of the final tone relative to the baseline. We derived the final height as follows. First, we calculated the speaker's base F0 range for a given sentence in Hertz from the maxF0 and minF0 of the sentence's base (e.g., 汪五 *Wang1 Wu3* 'Wang Wu'), using the formula in (1).

$$\text{speaker.F0-Range} = \text{base.maxF0} - \text{base.minF0} \quad (1)$$

Then, we calculated the maxF0 and minF0 of the final height from the maxF0s and minF0s of the sentence's base and final tone (e.g., T1 in 師 *shi1*), normalized in percentages of the speaker's base F0 range, using the formulae in (2-3).

$$\text{FH.maxF0} = ((\text{final.maxF0} - \text{base.maxF0}) / \text{speaker.F0-Range}) * 100 \quad (2)$$

$$\text{FH.minF0} = ((\text{final.minF0} - \text{base.minF0}) / \text{speaker.F0-Range}) * 100 \quad (3)$$

Due to articulatory inertia [11], typically the offset of a tone in one syllable is carried over to the onset of the next syllable [7, 8, 9]. In continuous speech, the maxF0 or minF0 of a syllable does not always reflect the maxF0 or minF0 of

the target tone on that syllable. For example, in Figure 4, *fen1* has a T1 [55] level tone, but the tone appears to be a T2 [35] rising tone due to the transition from a T3 syllable, *bai3*. In contrast, the sequence of T1s, *San1 jin1 tian1 ying1*, appears relatively level—with each tone following a T1—compared to the F0 contour of the rest of the sentence.

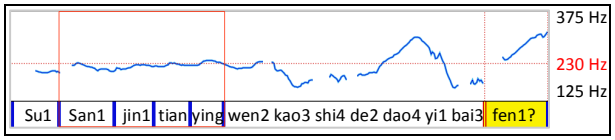

Figure 4: *'Su San got a hundred percent on her English test today?', produced by a female speaker.*

Therefore, our F0 analysis of the final height used either its maxF0 or minF0, whichever would be less likely affected by the preceding tone. We used a method similar to the ones used in [9, 12] to determine this value (which we refer to as "toneF0"), shown in (4).

If final tone = T1 or T4, toneF0 = FH.maxF0
If final tone = T2 or T3, toneF0 = FH.minF0 (4)

4. Results

4.1. Individual-Sentence Test

In this test, we compared statements with questions by calculating the overall mean toneF0 of the statements and the overall mean toneF0 of the questions. The boxplot in Figure 5 shows the mean toneF0 values, by sentence type and tone. For each of the four tones, the question's mean is higher than the statement's mean. In questions, the final T1 and T4's means are more than 50% above zero (the baseline), whereas the final T3's mean is more than 50% below zero. The final T2's mean is slightly above zero. In statements, all four final tones have negative means, which suggests final declination.

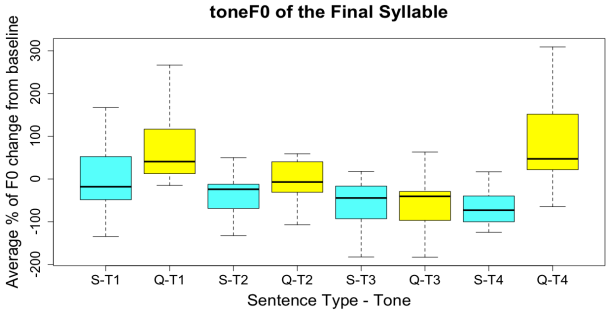

Figure 5: *The mean toneF0 values of statements (S) and questions (Q), separated into T1, T2, T3 and T4.*

A two-way ANOVA on toneF0 found main effects of sentence type [$F(1, 138) = 31.9$, $p < .001$] and final tone [$F(3, 138) = 8.0$, $p < .001$], as well as a significant interaction between sentence type and final tone [$F(3, 138) = 5.4$, $p = .002$], at $\alpha = .05$. Table 1 lists all of the significant post-hoc t-test results. These results indicate that 1) the speakers produced questions with a significantly higher toneF0 than statements when the final tone was T1 or T4; 2) in questions, the speakers produced T1 and T4 with significantly higher toneF0s than T3; and 3) in statements, the speakers produced T1 with a significantly higher toneF0 than T4.

Table 1: *Significant post-hoc test results of the interaction between sentence type and final tone, after Holm correction.*

Questions vs. Statements	Paired t-tests
Q:T1 – S:T1	[$t(22) = 4.3$, $p < .001$]
Q:T4 – S:T4	[$t(18) = 6.4$, $p < .001$]
Questions	**Two-sample t-tests**
Q:T1 – Q:T3	[$t(28) = 4.3$, $p < .001$]
Q:T4 – Q:T3	[$t(28) = 4.6$, $p < .001$]
Statements	**Two-sample t-tests**
S:T1 – S:T4	[$t(40) = 3.2$, $p = .003$]

Although the speakers produced all four final tones with higher toneF0s in questions than in statements, significant differences were found for T1 and T4 only. This result suggests that questions ending in T1 and T4 would be easier to identify than questions ending in T2 or T3. In addition, T4's significantly higher mean toneF0 in questions than in statements (as the example in Figure 6 shows) may explain why listeners in the experiment in [4] had an easier time identifying questions ending in T4 than in T2 or T3. Despite the lack of significant difference between T1 and T4 in questions, T1's mean toneF0 was significantly higher than T4's mean toneF0 in statements. The latter suggests a smaller F0 difference between statements and questions for T1 than for T4. Therefore, questions ending in T4 would be easier to identify than questions ending in T1. The paired-sentence test which follows seeks evidence to support this analysis.

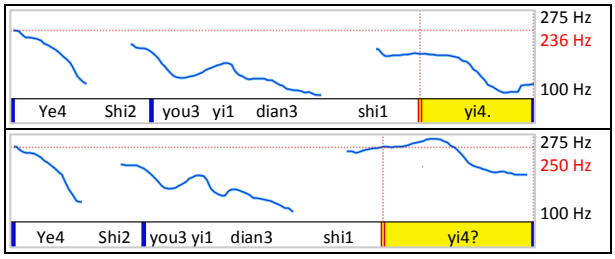

Figure 6: *'Ye Shi is slightly forgetful', produced by a male speaker.*

4.2. Paired-Sentence Test

In this test, we first calculated the difference in final height ("diffF0") between the sentences in each statement-question pair, using (5), and then compared the diffF0s across tones.

diffF0 = Q.final.toneF0 – S.final.toneF0 (5)

Figure 7 shows the mean diffF0s of all four tones.

A one-way ANOVA on diffF0 found a main effect of final tone [$F(3, 69) = 9.3$, $p < .001$]. Post-hoc Welch two-sample t-tests revealed that T4 had a significantly higher mean diffF0 than the other three tones: T1 [$t(31) = 3.2$, $p = .003$], T2 [$t(33) = 3.6$, $p = .001$] and T3 [$t(24) = 4.4$, $p < .001$]. This result indicates that, on average, the speakers produced the final T4 with the greatest difference in final height between statements and questions than any other tone. It supports our analysis that questions ending in T4 would be the easiest to identify. Indeed, the listeners in [4] identified questions ending in T4 more accurately than questions ending in the other tones. Since no other significant difference was found, we examined the F0 contours of the final tones for other potential cues for identifying echo questions ending in T2 or T3.

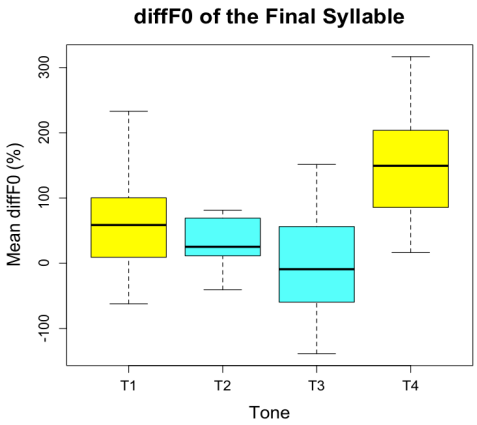

Figure 7: *The mean of the difference in F0 of a statement-question pair: T1 = 65.82, T2 = 51.93, T3 = -1.39 and T4 = 161.44.*

4.3. F0 Contours of the Final Tones

As the example in Figure 6 shows, the F0 contour of a final tone can differ when it appears in a question, rather than in a statement form. This variation could induce a timing difference in the F0 rise or fall between a question and its canonical statement. To capture the F0 contours of all four final tones, we used a Praat script to extract the F0 values at five different points of each tone: 0%, 25%, 50%, 75% and 100%, without normalization. Figure 8 shows the F0 contours of all four tones, approximated using the mean F0s at these five time points. Overall, questions had higher mean F0s than statements. [5] suggested an overall higher pitch for questions as well. However, the tonal patterns here reveal not only higher F0 values but also wider F0 ranges in questions than in statements [4]. As a result, the final F0 rise or fall is greater in questions than in statements. This pattern supports the claim that the F0 difference between statement and question intonations increases towards the end of the utterances [6]. Interestingly, for T4, the greatest F0 difference occurs at the 25% time point rather than at the end of the utterance.

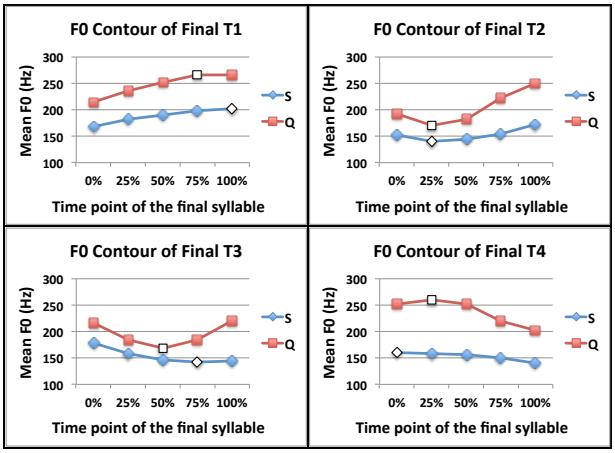

Figure 8: *Statement (S) and question (Q) F0 contours of all four final tones. The maxF0s of T1 and T4 and the minF0s of T2 and T3 are indicated in white.*

[13] found that the T3s in questions could reach the same low pitch level as the T3s in statements. In our study, the minF0s of the final T3 for both sentence types were also fairly close to each other, but so were the final T2's minF0s. However, T3 was realized as 21 in statement-final position, but as 214 in question-final position. The fact that T3's tail rises in questions but not in statements could be a cue for identifying questions ending in T3. Thus, it may explain why questions ending in T3 would be easier to identify than T2 [4]. In addition, the timing of T3's minF0 differed in questions and statements as a result of the difference in tonal patterns.

5. Discussion

Some researchers [14] claim that echo questions in Mandarin have a high boundary tone (H%) at the end of the utterance. Evidence from our study suggests a possible high tonal element that elevates the pitch of the final tone in echo questions. Its effect on T1 (a level tone) is a rise in pitch, while it increases the final rise in T2 (a rising tone) and T3 (a falling-rising tone), and it raises the onset of T4 (a falling tone), prior to the fall. These effects are different from the effects of an H% or a low boundary tone (L%) on echo questions in other languages. For example, in Cantonese, an H% triggers a steep F0 rise at the end of echo questions, regardless of the pitch level and direction of the final tone [15, 16, 17, 18]. In Mooré, a two-tone language, an L% triggers an F0 fall at the end of echo questions and lengthens the final vowel [19]. In both of these languages, the boundary tones occur after the final tone. In our study, however, the gradual rise in pitch towards the end of the echo questions is similar to the effect on yes/no questions in [20]. The timing of the rise on the final tone is aligned with the high pitch level of the tone (e.g., on pitch level 5 of T4 [51]). Therefore, it is uncertain whether an H% triggered the rise on the final tone of the echo questions here.

The F0 contours in Figure 8 suggest that if listeners (e.g., in [4]) were to focus on the F0 difference between statements and echo questions at the end of the contour, they would likely identify questions ending in T2 or T3 easier than T1 or T4. At the 100% time point, the F0 differences for T2 (77 Hz) and T3 (76 Hz) are greater than the F0 differences for T1 (64 Hz) and T4 (63 Hz). However, if they were to focus on the F0 difference at the turning point of the tonal contour (e.g., the transition point between pitch levels 3 and 5 for T2 [35]), they would have a harder time identifying questions ending in T2.

6. Conclusion

This study has demonstrated that the F0 height of the final tone is a potential cue for the identification of echo questions versus statements in Mandarin. However, since the speakers in this study produced this cue differently across all four tones, this cue must work in conjunction with other prosodic cues to signal the question/statement distinction. For example, since there is no significant difference in F0 minimum between statements and echo questions ending in T3, speakers might use a tail rise to indicate questions, or creakiness to indicate statements. Also, declination is more evident in statements than in echo questions in Mandarin, so declination could serve as a cue for identifying statements.

In future research, we aim to re-run the tests on a larger set of data and also include other factors, such as the duration of the final tone and the extent of the F0 rise or fall. We would also like to follow up with a perception study to find out if listeners use the final F0-height cue for the identification of echo questions and if the timing of this cue matters.

7. Acknowledgements

We sincerely thank the reviewers and the audience of the 2016 TAL for their comments. This research was supported by the Social Sciences and Humanities Research Council of Canada.

8. References

[1] Li, C. N. and Thompson, S. A., Mandarin Chinese: A Functional Reference Grammar, University of California Press, 1981.

[2] Chao, Y.-R., Mandarin Primer: An Intensive Course in Spoken Chinese, Harvard University Press, 1948.

[3] Yuan, J., "Perception of intonation in Mandarin Chinese", J. Acoust. Soc. America, 130(6):4063-4069, 2011.

[4] Yuan, J. and Shih, C., "Confusability of Chinese intonation", Proc. Speech Prosody, Nara, Japan, 2004.

[5] Yuan, J., Shih, C. and Kochanski, G. P., "Comparison of declarative and interrogative intonation in Chinese", Proc. Speech Prosody, Aix-en-Provence, France, 2002.

[6] Liu, F., Surendran, D. and Xu, Y., "Classification of statement and question intonations in Mandarin", Proc. Speech Prosody, Dresden, Germany, 2006.

[7] Gandour, J., Potisuk, S. and Dechongkit, S., "Tonal coarticulation in Thai", J. Phonetics, 22:477-492, 1994.

[8] Xu, Y., "Contextual tonal variations in Mandarin", J. Phonetics, 25:61-83, 1997.

[9] Wang, B. and Xu, Y., "Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese", J. Phonetics, 39:595-611, 2011.

[10] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer", Computer application, version 5.3.51. Online: http://www.praat.org, accessed on 30 May 2013.

[11] Cheng, C. and Xu, Y. "Articulatory limit and extreme segmental reduction in Taiwan Mandarin", J. Acoust. Soc. America, 134(6):4481-4495, 2013.

[12] Chen, Y. Y. and Gussenhoven, C., "Emphasis and tonal implementation in Standard Chinese", J. Phonetics, 36:724-746, 2008.

[13] Yuan, J., "Mechanisms of question intonation in Mandarin", in Q. Huo, B. Ma, E.-S. Chang and H. Li [Ed], Chinese Spoken Language Processing, 19-30, Springer, 2006.

[14] Peng, S.-H., Chan, M. K. M., Tseng, C.-Y., Huang, T., Lee, O. J. and Beckman, M. E., "Towards a Pan-Mandarin system for prosodic transcription", in S.-A. Jun [Ed], Prosodic Typology: The Phonology of Intonation and Phrasing, 230-270, Oxford University Press, New York, 2005.

[15] Wong, W. Y. P., Chan, M. K. M. and Beckman, M. E., "An autosegmental-metrical analysis and prosodic annotation conventions for Cantonese", in S.-A. Jun [Ed], Prosodic Typology: The Phonology of Intonation and Phrasing, 271-300, Oxford University Press, New York, 2005.

[16] Gu, W., Hirose, K. and Fujisaki, H., "Analysis of the effects of word emphasis and echo questions on F0 contours of Cantonese utterances", Proc. Interspeech, 1825-1828, Lisbon, Portugal, 2005.

[17] Ma, J. K., Ciocca, V. and Whitehill, T. L., "Effect of intonation on Cantonese lexical tones", J. Acoust. Soc. America, 120(6): 3978-3987, 2006.

[18] Ma, J. K.-Y., Ciocca, V. and Whitehill, T. L., "The perception of intonation questions and statements in Cantonese", J. Acoust. Soc. America, 129(2):1012-1023, 2011.

[19] Rialland, A., "The African lax question prosody: Its realisation and geographical distribution", Lingua, 119(6):928-949, 2009.

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