

TONE, INTONATION, AND EMPHATIC STRESS IN L2 MANDARIN SPEECH BY ENGLISH AND CANTONESE LEARNERS

Wentao Gu

Nanjing Normal University, China
wtgu@njnu.edu.cn

ABSTRACT

On the basis of a set of well-controlled sentences varying in sentence type, tone identity, and focus position, the present study compared F_0 characteristics of Mandarin speech between native speakers and two groups of L2 learners whose native languages were Cantonese and English, respectively. The results showed systematically that most L2 errors in the F_0 manifestations of tone, intonation, and emphatic stress could be explained by language transfer effects. The findings can be used to guide L2 learning of Mandarin.

Keywords: tone, intonation, emphatic stress, Mandarin, L2 error.

1. INTRODUCTION

The naturalness of speech depends highly on its prosody consisting of the components such as tone, intonation, and stress (lexical or emphatic stress). The acoustic manifestations of these prosodic components as well as the manner how they interact with each other can be quite language-dependent. A number of previous studies (e.g., [8]) have shown that many L2 prosodic errors are attributed to language transfer effects resulting from these cross-linguistic differences.

The present work inspects F_0 errors in L2 Mandarin. Two other languages concerned here are standard Cantonese spoken in Hong Kong and Macau (hereafter Cantonese) and American English (hereafter English). These three languages contrast sharply in prosodic phonology in terms of tone, intonation, and stress.

Mandarin and Cantonese are both Chinese tone languages, in which a syllable is divided into an initial consonant and a final rhyme carrying a tone which is cued mainly by F_0 . 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 falling-rising, or ‘dipping’ tone) while Cantonese has six lexical tones. Cantonese has no contrast in lexical stress, while Mandarin does – there is no lexical stress in the phonological sense but a neutral tone T0 exists, functioning as unstressed. English, on the contrary, is not a tone language but a stress language, in which the syllables in a word differ in the degree of lexical stress.

In all three languages, statements and unmarked yes-no questions (intonation question) are associated with falling and rising sentence intonation, respectively. Their patterns of F_0 raising in questions, however, are

different. While Cantonese and English raise F_0 almost at the end of a question [4, 6, 9] – hence a boundary tone is introduced in the AM theory, Mandarin raises F_0 in a global domain (viz., questions start at a higher F_0 than statements) [7, 12] – though F_0 rising may still peak at the end [5]. For emphatic stress, which is associated with focus, previous studies have shown language-dependent effects on F_0 : asymmetric for Mandarin and English, i.e., on-focus raising/expansion and post-focus suppression (i.e., lowering/compression) of F_0 ranges [10, 11], while symmetric (no post-focus suppression) for Cantonese [1, 2]. Also, the interaction between question and emphasis has been studied [5].

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 used read speech of a set of controlled sentences, for which the Chinese text, the pinyin transcription, and the English translation are shown below:

他帮[朱/竺/祖/祝]科长[发/砸/打/炸][包/桃/枣/炮]./?

Pinyin: “Ta1 bang1 [zhu1/zhu2/zu3/zhu4] ke1zhang3 [fa1/za2/da3/zha4] [bao1/tao2/zao3/pao4]./?”

English: He’ll help manager [Zhu/Zu] [verb] [noun]./?

These sentences share a fixed carrier frame, in which three target syllables shown in brackets are embedded: a surname at the sentence-medial position, a verb at the penultimate position, and a noun at the final position. The target syllable at each position has four options, sharing the same rhyme (/u/, /a/, /ao/, respectively) but varying in tones. The initial consonants of all syllables are voiceless. As tone is carried mostly by the rhyme of a syllable, this design can minimize segmental effects on tones. All these sentences are meaningful texts, varying systematically in three factors:

- Sentence type (2 levels): statement vs. unmarked yes-no question (intonation question).

- Lexical tones in three target syllables (4 levels for each): T1~T4; neutral tone was not included.
- Position of focus (4 levels): broad focus (i.e., neutral focus) and three conditions of narrow focus (i.e., on the sentence-medial, -penultimate, and -final target syllables, respectively).

Thus, there are altogether $2 \times 4 \times 4^3 = 512$ sentences.

Three groups of subjects participated in the experiment: MS, CS, and ES, who were native in Mandarin, Cantonese (HK or Macau), and American English, respectively. Each group consisted of six subjects (3M+3F) at similar ages – the average ages for the three groups are 25, 19, and 24, respectively. The native subjects in MS were graduate students with high Mandarin proficiency, while those in CS and ES were L2 Mandarin learners at the intermediate level – they had studied Mandarin for 1~2 years and had already passed Level 5 of HSK, the Chinese proficiency test.

In recording, the sentences with a broad focus were read aloud by the subjects (S) directly, whereas those with a narrow focus were elicited by a prompt utterance from the experimenter (E). For example, the following two dialogues elicited a narrow focus on 朱 in statement and question, respectively, by the subjects.

- E: 他帮李科长发包吗?
(Will he help manager Li distribute the bags?)
S: 他帮朱科长发包。
(He will help manager Zhu distribute the bags.)
- E: 他不是帮李科长, 而是帮朱科长发包。
(He will not help manager Li but manager Zhu to distribute the bags.)
S: 他帮朱科长发包?
(He will help manager Zhu distribute the bags?)

Speech recording was conducted in a sound-proof room after the subjects got familiar with the reading materials and made sure to know the exact pronunciations of all the texts. The recording was done at the normal speech rate of each subject, and was monitored by the experimenter. Once there was a mispronunciation or disfluency, the subject would be asked to repeat recording that sentence until success.

Segmentation of utterances into syllable initials and rhymes was done manually. F_0 values were extracted at 10ms intervals using autocorrelation analysis in Praat. After manual correction of gross errors in F_0 extraction, the raw F_0 values were smoothed, and also interpolated if there are interruptions within the rhymes. Ignoring durational differences, the syllable rhyme based time-normalized F_0 contours were obtained by extracting the F_0 values at 10 equally-spaced points in the rhyme of each syllable, and then were averaged among a certain set of subjects and utterances in the scale of semitone.

3. RESULTS

Figures 1 and 2 show the average time-normalized F_0 contours measured in semitone, varying in the tones of

the sentence-medial and -final target syllables, respectively. Namely, in each figure, the F_0 contours in the other two target syllables are averaged over all four tones. The left and right columns of panels in each figure refer to two focus conditions, i.e., broad focus and narrow focus on the corresponding target syllable. The solid and dashed lines indicate F_0 contours in statements and questions, respectively. Four colours are used to indicate the sentences with four different tones in the corresponding target syllable.

Based on Figs. 1–2, the F_0 errors in L2 Mandarin by the two L2 groups can be found by comparison with the native MS. Due to space limitations, in the following we will focus on qualitative descriptions without giving all details of statistical analyses.

3.1. Tone

The F_0 contours in the target syllables show that the two L2 groups largely grasped the basic F_0 patterns of Mandarin four tones, i.e., T1 high, T2 rising, T3 low or dipping (at non-final or final positions, respectively), and T4 falling. The most obvious L2 error in tones, as also observed in previous studies, lies in F_0 range – CS and MS have narrower F_0 ranges than MS, especially for the two contour tones T2 (rising) and T4 (falling).

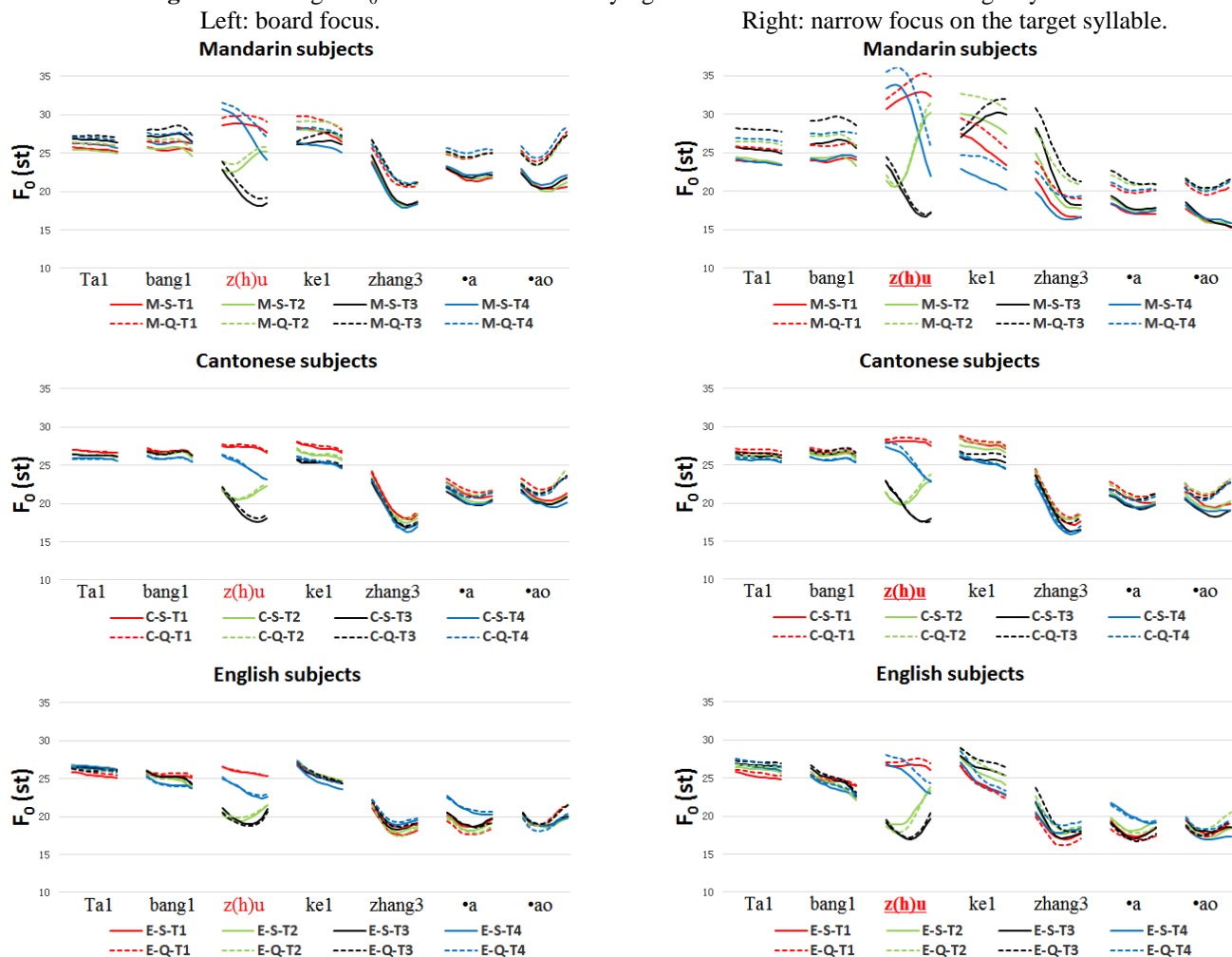
In addition, there is an important L2 error in F_0 shape – the three groups produced T3 differently. In the sentence-medial T3 syllable, ES produced a dipping F_0 contour, while the other two groups did not produce a rising tail. Even in the sentence-final T3 syllable where MS produced a dipping contour, ES showed a still higher rising tail in statement – the falling and rising parts have almost equal amplitudes. This is a kind of over-articulation for non-final T3, quite possibly because English learners are usually taught T3 in its isolated form without being aware that the rising tail should be ignored for non-final T3, and they might be excessively cautious of producing tones of each syllable even in continuous speech.

In the sentence-final T3 syllable, it should also be noted that CS produced a lower rising tail than MS. This can be explained by the fact that CS's native tone language Cantonese does not have a dipping tone.

3.2. Intonation

For statements, by comparing the several T1 syllables in the broad focus condition, we can observe that the F_0 declination effect is more distinct for L2 groups, especially for ES – this may be because native speakers tend to be more flexible in phrasing or accentuation by resetting F_0 . The most conspicuous L2 error in sentence intonation, however, is on questions. It is known that intonation question generally has higher F_0 than statement, but the details of F_0 raising can be language-dependent. Table 2 shows the average F_0 increments (in semitone) from statement to question in each syllable, with statistical significances indicated.

Figure 1: Averaged F_0 contours with four varying tones in the sentence-medial target syllable.



For MS, in the broad focus condition, F_0 raising occurs over the entire utterance, while the amplitude of F_0 raising increases with the time course – the later the larger, jumping to the peak in the sentence-final syllable. This kind of global F_0 raising is even more conspicuous in narrow focus conditions, where F_0 is raised relatively high even at the beginning of the utterance (hence the gradual increase of F_0 raising are not valid) and finally it jumps to the peak at the end of the utterance.

In contrast, the two L2 groups' F_0 raising occurs much later and is much weaker in all focus conditions. ES does not show any significant and meaningful F_0 difference between question and statement. For CS, there is significant F_0 raising (but much weaker than MS) which however is highly localized at the utterance ending, mostly in the *latter* half of the final syllable.

These L2 errors in intonation can be explained by language transfer effects. As in their native languages both Cantonese and English implement questions with a localized instead of a global F_0 raising, they tend to maintain this strategy in their L2 Mandarin. Moreover, unlike other two languages, Cantonese has very rich sentence-final particles, which contribute substantially to Cantonese intonation. So, it is natural for CS to raise F_0 relatively strongly in the final syllable in L2 speech.

Right: narrow focus on the target syllable.

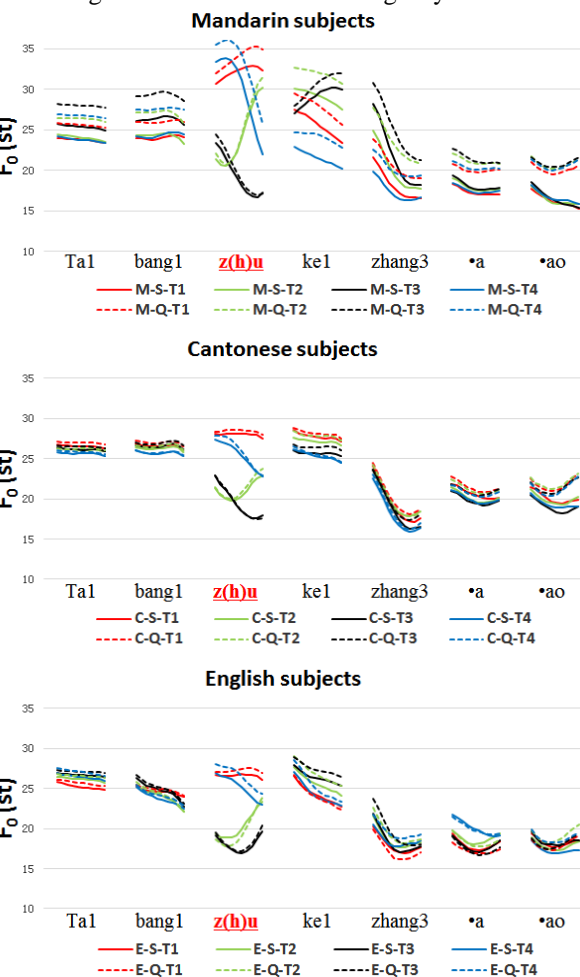


Table 2: F_0 increments from statement to question in all syllables of the sentence (in semitone).

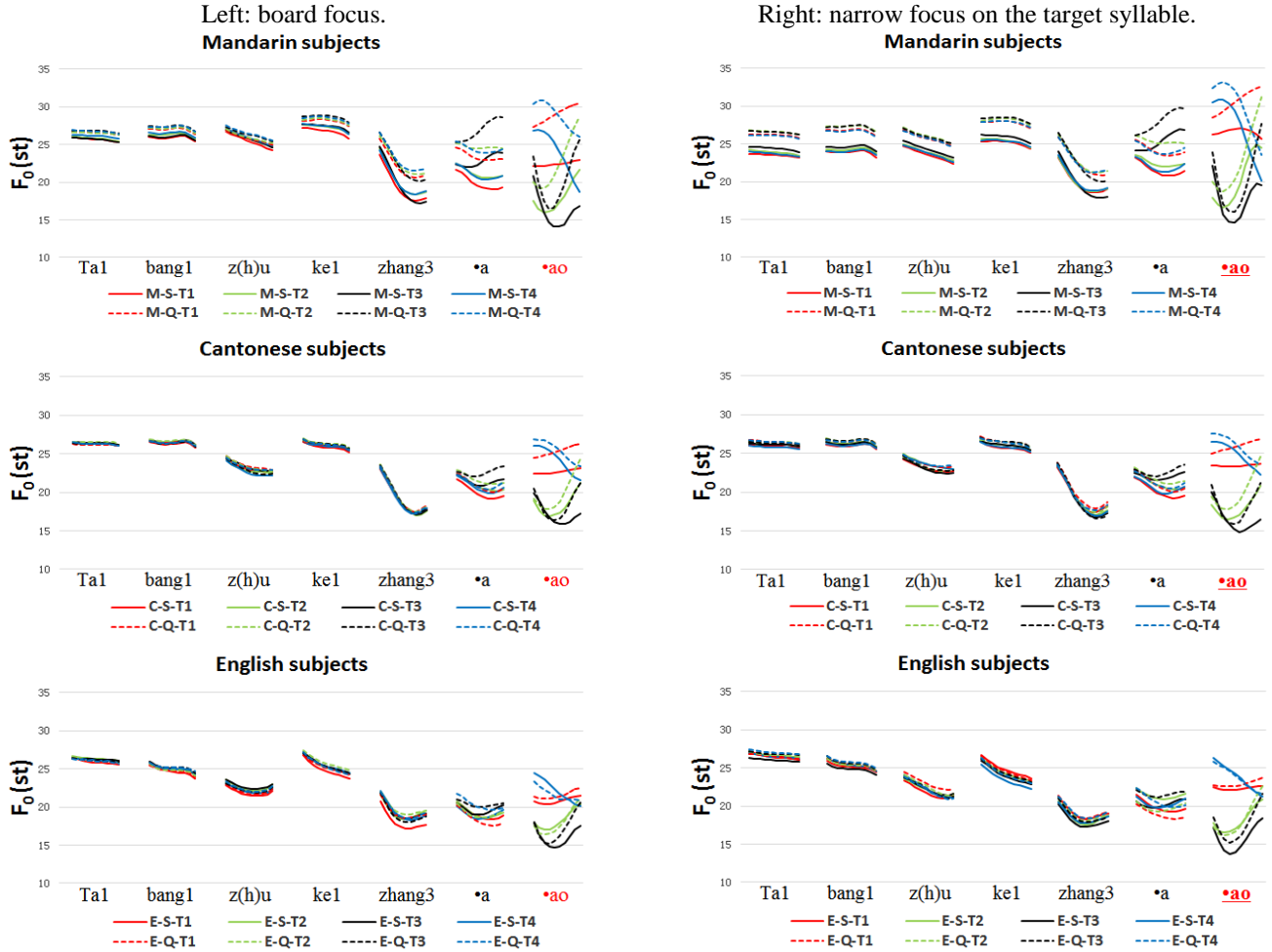
Group	Focus	ta1	bang1	z(h)u	ke1	zhang3	*a	*ao
MS	broad	0.7 [†]	1.0 [†]	1.2 [*]	1.4 [*]	2.6 [*]	2.7 [*]	4.1 [*]
	n-medial	2.4 [†]	2.7 [†]	1.5 [*]	2.3 [*]	3.0 [*]	3.0 [†]	4.1 [†]
	n-penult	2.8 [†]	2.9 [†]	2.0 [†]	2.6 [†]	2.4 [†]	2.2 [*]	3.9 [*]
	n-final	2.5 [†]	2.8 [†]	2.0 [†]	2.6 [†]	2.6 [†]	2.6 [*]	3.1 [*]
CS	broad	0.0	0.0	0.3	0.2	0.2	0.8 [*]	1.7 [*]
	n-medial	0.3 [*]	0.3 [*]	0.3 [†]	0.5 [*]	0.5 [†]	0.9 [*]	2.0 [*]
	n-penult	0.5 [*]	0.5 [*]	0.5	0.6 [*]	0.4 [*]	0.7 [*]	1.7 [*]
	n-final	0.5 [*]	0.5 [*]	0.1	0.5 [*]	0.3	0.7 [†]	1.8 [*]
ES	broad	-0.1	0.0	0.0	0.3	0.3	-0.3	0.3
	n-medial	0.5 [†]	0.3	0.4	0.8 [†]	0.6	-0.4	0.6
	n-penult	-0.1	-0.0	0.4	0.2 [†]	0.4	0.6 [†]	1.1
	n-final	0.4	0.4	0.3	0.2	0.5	-0.2	0.6

* $p < 0.01$; † $0.01 < p < 0.05$.

3.3. Emphatic stress

In line with previous findings, for MS, emphasis leads to both on-focus F_0 raising/expansion and post-focus F_0 lowering/compression (the post-focus lowering here is most prominent in the final two syllables, for /z(h)u/ and /ke1zhang3/ constitutes one word – hence the latter might also be affected when the former is emphasized). In contrast, ES shows similar but weaker effects, while

Figure 2: Averaged F_0 contours with four varying tones in the sentence-final target syllable.



CS shows even weaker effects on F_0 . This coincides with the different natures of their native languages, i.e., English has similar effects of emphatic stress on F_0 as Mandarin [10, 11], while Cantonese does not [1, 2].

3.4. Interaction effects

Tone, intonation and emphasis are simultaneously transmitted in the F_0 contour of an utterance. There can be complex interactions between the three factors.

The interaction between tone and intonation is a critical issue in tone languages. The most conspicuous interaction lies in the final syllable of a question with a rising intonation. For MS, question intonation does not modify the relative tone patterns in the final syllable but raises F_0 of the entire syllable – the later the larger raising. Thus, the so-called ‘boundary tone’ here is neither a tone replacement nor an even addition. By comparison, CS shows quite similar interaction effects as MS, most possibly due to their common grounds as tone languages. For ES, there is no explicit effect of question intonation, not to mention any interaction.

Another interaction is shown between emphasis and question intonation. When there is a narrow focus in a question, the F_0 raising relative to statement tends to be earlier and higher – neither starts from nor peaks at the on-focus syllable, which disagrees with the boosting

effect on the focused syllable as reported in [5]. For CS, the observed interaction effect is weaker than MS.

Finally, the interaction between tone and emphasis is consistent among three groups, i.e., emphasis does not alter qualitative tone patterns, but affects high pitch targets more than low ones. This is because high pitch is easier to manipulate for physiological reasons, which coincides with the findings across languages [2, 3, 10].

4. CONCLUSION

Based on a set of well-controlled Mandarin sentences, we examined the F_0 manifestations for tone, intonation, and emphasis as well as the interaction effects between them, for three groups whose native languages contrast in prosodic phonology on these three factors. The study found that most L2 errors on F_0 were related to language transfer effects. Also, teaching methods play a crucial role – some L2 errors may be easily overcome if the learners know the rules, e.g., two variants of T3 at different positions. In any case, awareness of the common prosodic error patterns for a particular L1 group should be helpful for L2 teaching and learning.

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

5. REFERENCES

- [1] Gu, W., Hirose, K., Fujisaki, H. 2006. Modeling the effects of emphasis and question on fundamental frequency contours of Cantonese utterances. *IEEE Trans. Audio, Speech & Lang. Proc.* 14, 1155–1170.
- [2] Gu, W., Lee, T. 2009. Effects of tone and emphatic focus on F_0 contours of Cantonese speech: A comparison with Standard Chinese. *Chinese J. Phonetics* 2, 133–147.
- [3] Jia, Y., Xiong, Z., Li, A. 2006. Phonetic and phonological analysis of focal accents of disyllabic words in Standard Chinese. *Proc. 6th ISCSLP* Singapore, 43–54.
- [4] Ladd, D. R. 1996. *Intonational Phonology*, Cambridge: Cambridge University Press.
- [5] Liu, F., Xu, Y. 2005. Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica* 62, 70–87.
- [6] Ma, J. K., Ciocca V., Whitehill T. L. 2006. Effect of intonation on Cantonese lexical tones. *JASA* 120, 3978–3987.
- [7] Shen, X. N. 1990. *The Prosody of Mandarin Chinese*, Berkeley: University of California Press.
- [8] Trouvain, J., Gut, U. (eds). 2007. *Non-Native Prosody: Phonetic Description and Teaching Practice*. Berlin: Mouton de Gruyter.
- [9] Xu, B. R., Mok, P. 2011. Final rising and global raising in Cantonese intonation. *Proc. 17th ICPhS* Hong Kong, 2173–2176.
- [10] Xu, Y. 1999. Effects of tone and focus on the formation and alignment of f_0 contours. *J. Phonetics* 27, 55–105.
- [11] Xu, Y., Xu, C. X. 2005. Phonetic realization of focus in English declarative intonation. *J. Phonetics* 33, 159–197.
- [12] Yuan, J. 2006. Mechanisms of question intonation in Mandarin, *Proc. 6th ISCSLP* Singapore, 19–30.