CORRELATES OF CHINESE NEUTRAL TONE PERCEPTION IN DIFFERENT CONTEXTS

Aijun Li, Shanshan Fan

Institute of Linguistics, Chinese Academy of Social Sciences
liaj@cass.org.cn, fanshan08@163.com

ABSTRACT

A psychoacoustic experiment was conducted to investigate correlates of neutral tone perception and the distribution of the perceptual space in three contexts, i.e. isolation, on-focus and post-focus for Standard Chinese. To this end, we adopted a minimal word pair, i.e. /ʂɤ^{2}tʰəu^{2}/(“蛇头”, head of snake) and /ʂɤ^{2}tʰəu^{0}/ (“舌头”, tongue) where tonal contrast is between a neutral tone T0 and a rising tone T2 (normal stressed tone). The result shows that pitch is always a more reliable cue than duration in neutral tone perception. However, the contribution of different acoustic cues and the distribution of the perceptual spaces are not only related to contexts, but also to the differences between the tonal spaces of contrastive tones. The way acoustic cues influence the perception of unstressed items in Chinese is not always consistent with that in stress languages such as English and Dutch.

Keywords: neutral tone perception, Standard Chinese, acoustic cues, contexts, stress

1. INTRODUCTION

In Standard Chinese, besides the four distinctive lexical tones, there exist weak syllables in terms of neutral tone [1, 2]. Neutral tone can be considered to be related to both tone and stress system in Standard Chinese [4]. It doesn’t occur in the initial position of a word and is assumed to be associated with weak syllable, which is short and light, such as the second syllable in /ti^{4}ti^{0}/ (“弟弟”, little brother). Hereafter, “1~4” stands for four lexical tones (T1~T4 in short) and “0” for neutral tone (T0). Neutral tone has a mid-pitch target and its F_0 realization is associated with the preceding tone [3]. Morphologically, some neutral tone words have contrastive meanings with their normal stressed counterparts. For example, with normal stress (strong-strong), the word “地道”/ti^{4}tao^{4}/ means ‘tunnel’, while with neutral tone (strong-weak) , the word “地道”/ti^{4}tao^{0}/ means ‘authentic’. Studies on the acoustic correlates of neutral tone and its perceptual space will help understand the characteristics of the spoken Chinese.

Previous experimental studies confirmed that the relevant acoustic cues for perceiving stress include pitch, duration, intensity, spectral balance or spectral tilt (timber). In stress languages such as English and Dutch, pitch is the most salient acoustic cue [5] for pitch accents at the utterance level [6, 7, 8]. But for lexical stress, duration is the most reliable cue for Dutch listeners. Spectral balance or spectral tilt is an important cue, but not as reliable as duration. Overall, intensity is the least important cue for stress perception in Dutch and English. In English, vowel reduction is a pervasive phenomenon in unstressed syllables. However, vowel reduction is less pervasive and it is the poorest cue in Dutch [7, 9].

In Standard Chinese, the acoustic correlates of word or sentence stress are identical with those in stress languages. However, in regard to pitch and duration, which one is more relevant to the weak syllable remains an issue of debate. Some researches indicated that pitch outweighs duration [10, 11], whereas others reported the opposite result [12]. In terms of duration, the unstressed syllable in neutral tone shrinks to 50% [13~17] or 60% [3] of its stressed counterpart. On average, the unstressed syllable is 60% of the preceding stressed syllable [10]. Regarding intensity, the unstressed syllable is not necessarily lighter than the stressed one. Therefore, intensity does not seem a reliable cue for neutral tone perception [10, 12, 14, 18]. Spectral tilt may have great influence on neutral tone perception, but it is less important than duration [19].

To explore the contribution of pitch and duration to neutral tone perception, a psychoacoustic experiment was conducted in [23] for a minimal stressed pair T1-T0 in three contexts, isolation, on-focus and post-focus. It is found that pitch is always a more reliable cue than duration, and the influence of pitch and duration is closely related to the context. In the present study, we follow up the experiment in [23] to check another minimal pair of T2-T0 by using word pair /ʂɤ^{2}tʰəu^{2}/ and / ʂɤ^{2}tʰəu^{0}/. The tonal space between contrast T2 (LH) and T0 (L) is greater than that between T1 (HH) and T0(L) used in [23]. Please refer Figure 1 to understand the meaning of tonal space which is the acoustic space between normal stress tone and neutral tone. The present study will focus on the correlates of neutral tone perception by checking: 1) the contribution of pitch and duration to neutral tone perception relating to context and tonal space, and 2) the distribution of perceptual space along dimensions of pitch and duration relating to context and tonal space.

2. PERCEPTUAL EXPERIMENT

We follow the procedure and setup of the perceptual experiment on the minimal pair of words /ʂɤ^{2}t^{h}əu^{2}/ and /ʂɤ^{2}t^{h}əu^{0}/ as originally presented in [23].

Target words were produced by a male standard Chinese speaker in three contexts:

(1) **Isolation**: /ʂɤ^{2}t^{h}əu^{2}/ and /ʂɤ^{2}t^{h}əu^{0}/;

(2) **On-focus**: target words were embedded in a carrier sentence in the focus position:

Elicited sentence: Which word did Xiaozhao learn yesterday?

Carrier sentence: Yesterday Xiaozhao learned **/ʂɤ^{2}t^{h}əu^{2}/ (/ʂɤ^{2}t^{h}əu^{0}/)** this word.

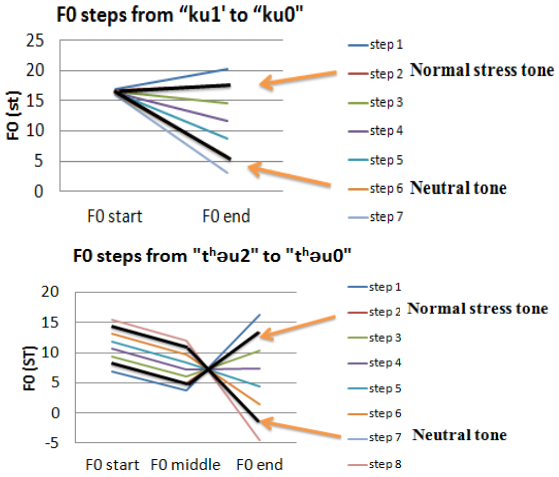
(3) **Post-focus**: target words were embedded in a carrier sentence in the post-focus position:

Elicited sentence: When Xiaozhao learn the word /ʂɤ^{2}t^{h}ou^{2}/ (/ʂɤ^{2}t^{h}əu^{0}/)?

Carrier sentence: Xiaozhao **yesterday** learned ʂɤ^{2}t^{h}əu^{2}/ (/ʂɤ^{2}t^{h}əu^{0}/) this word.

PRAAT (http://www.fon.hum.uva.nl/praat/) was applied to extract F_0 and duration of the target words.

Figure 1: F0 steps for T1-T0 [23] and T2-T0



The stimuli were achieved through systematical manipulation on the variation of pitch (P) and duration (D) between the neutral tone syllables and their stressed counterparts. One extra curve was added above the original F_0 curve of /t^{h}əu^{2}/ and one below that of /t^{h}əu^{0}/ with the same step (Figure 1). Two more steps of duration were added longer than /t^{h}əu^{2}/ and shorter than / t^{h}əu^{0}/.

Table 1: Pitch and duration in varying steps and the number of target stimuli in three contexts

Contexts	P steps	D steps	Stimuli number
isolation	8	13	8*13=104
on-focus	8	11	8*11=88
post-focus	7	10	7*10=70
Total			262

Pitch and duration in varying steps as well as the number of target stimuli in different contexts are listed in Table 1. Please note that the stimuli of on-focus and post-focus contexts included the whole carrier utterances rather than the target words only. Besides, 82 fillers (including 27 words and 55 utterances) were added. Therefore, the total number of stimuli finally used in the experiment was 344.

Sixteen subjects, including 8 males and 8 females, made their 3-way judgments on words in neutral tone /ʂɤ^{2}t^{h}əu^{0}/ or word carrying normal stress /ʂɤ^{2}t^{h}əu^{2}/ or uncertainty. All stimuli were played by E-prime randomly.

3. ANOVA ANALYSIS

The perceptual score was set to 3,1 and 2 points for normal stressed tone, neutral tone and uncertainty respectively. ANOVA was employed to analyze the perception results in three contexts, with pitch and duration as independent variables, the perceptual score as dependent variable, and subjects as co-variable.

The ANOVA results in Table 2 demonstrated the following. (i) Pitch significantly affects neutral tone perception across all contexts ($P<0.05$). (ii) Duration is reliable in isolation, but not in either on-focus or post-focus position ($P_{on\text{-}focus}=0.629>0.05$; $P_{post\text{-}focus}=0.139>0.05$). (iii) Comparing F-values in three contexts, the amplitude of pitch influence in descending order is isolation> on-focus > post-focus. (iv) The interaction of pitch and duration (P * D) is not significant.

Table 2: ANOVA analysis on perceptual results

Contexts	Variances	df	F	Sig.
isolation	P	7	826.461	.000
	D	12	2.956	.000
	P * D	84	.933	.0649
on-focus	P	7	225.603	.000
	D	10	.800	.629
	P * D	70	0.810	.869
post-focus	P	6	54.924	.000
	D	9	1.510	.139
	P * D	54	1.003	.470

Figure 2 plots the average perceptual scores (vertical axis), and variation of different acoustic correlates (horizontal axis). The first column displays the average perceptual scores as a function of pitch and duration. The second and the third columns are plots of the average perceptual scores as a function of pitch and duration respectively.

In three contexts, pitch plays a significant role in perception as shown in the first and second columns. As the increase of pitch steps, shifting from T2 to T0, the perceptual scores decrease from 3 to 1. The perceptual curves for pitch tend to be more similar to a typical function of categorical perception in focus and on-focus contexts (mid column).

The duration shows some relation to the perceptual score in isolation, but little relation in other two conditions (refer to the last columns).

The results demonstrated that, for T2-T0 contrast, the contribution of acoustic cues to neutral tone perception is different across the three contexts. When comparing the results for T1-T0 contrast in [23], we found that contribution of acoustic cues to neutral tone perception is distinct between these two contrasts. For example, duration of T1-T0 contrast has significant effects across all contexts, and even in post-focus it has an almost similar effect as pitch. Besides, the perceptual patterns between these two contrasts are also inconsistent.

Till now both statistical results and average perceptual patterns point to the fact that in addition to contexts, tonal space could be another correlate of neutral tone perception.

Figure 2: Average perceptual results in three contexts (vertical axis stands for the perceptual score)

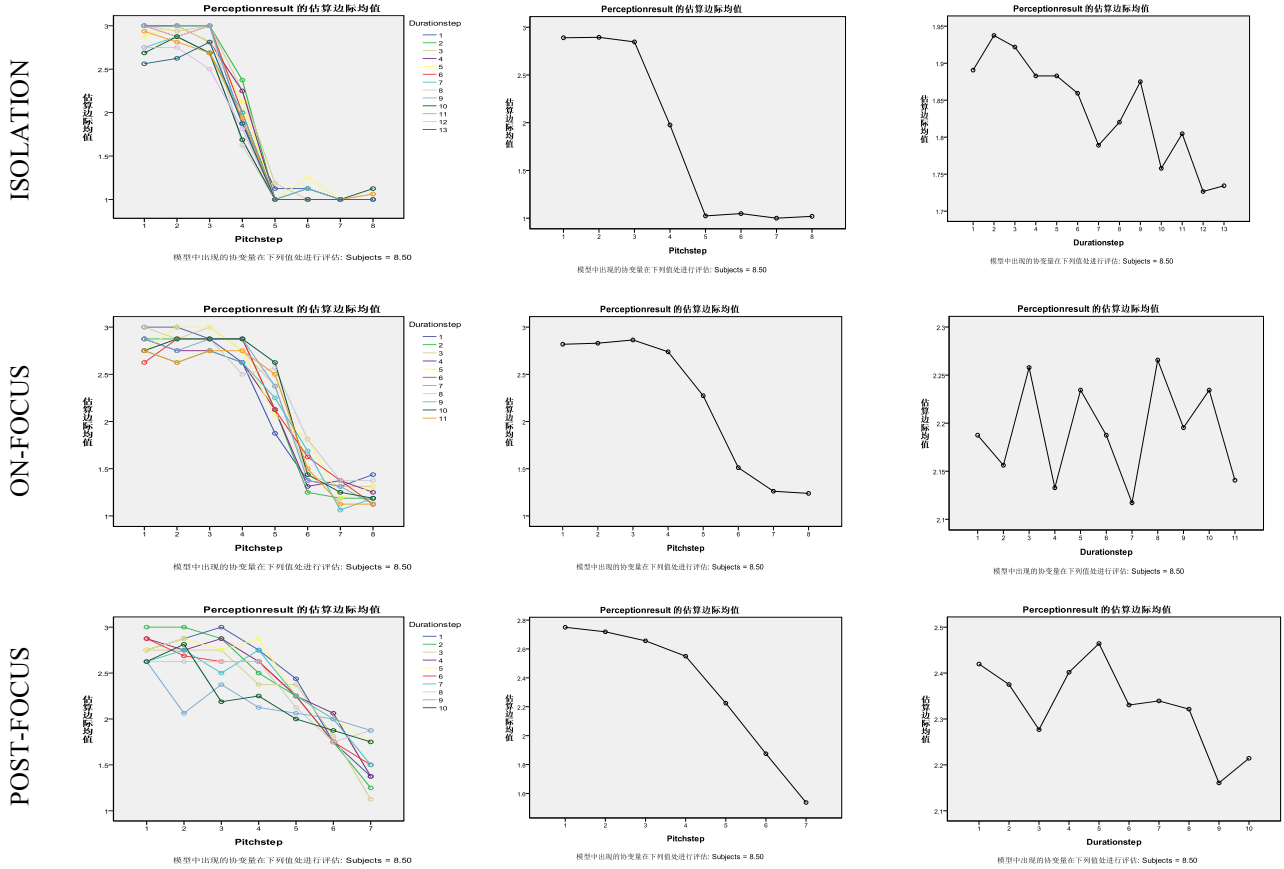
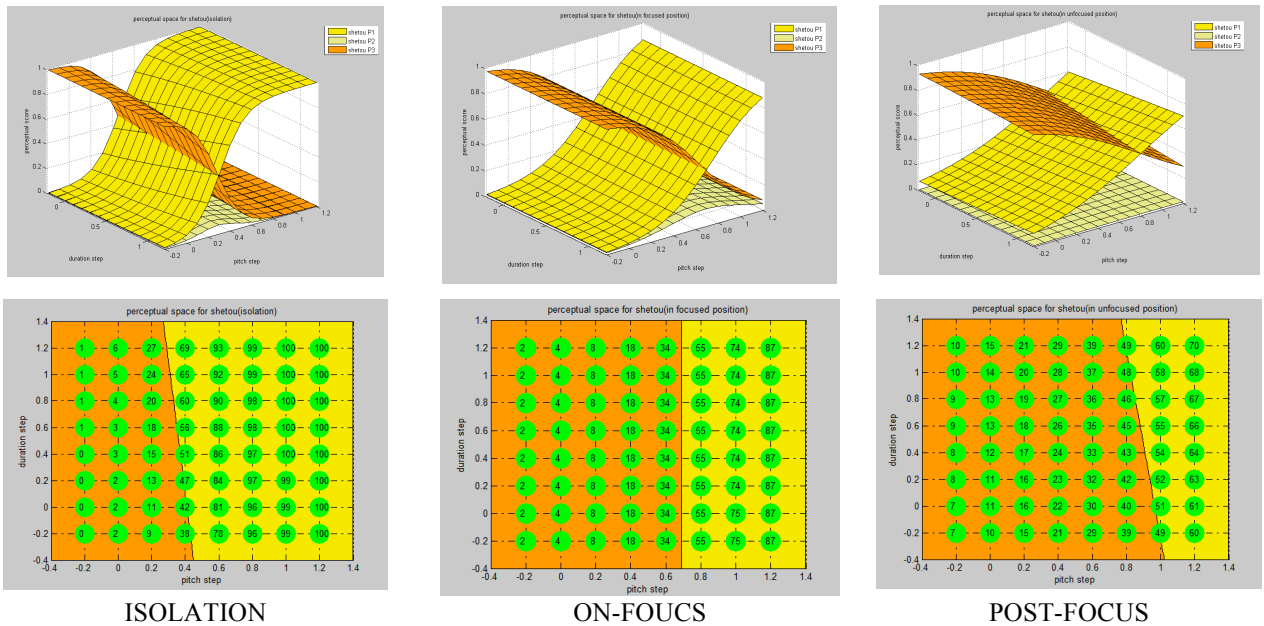


Figure 3: The 3D and 2D simulated spaces, P1 shown in yellow or light and P3 in orange or dark



ISOLATION ON-FOUCS POST-FOCUS

4. SIMULATION OF THE PERCEPTUAL SPACES

To clearly display and compare the perceptual spaces between the normal stressed tone and neutral tone in three contexts, a Multinomial Ordinal Logistic Regression Analysis was conducted to simulate the perceptual results, following the same method as described in [23]. Independent variables were pitch and duration. Nine logistic functions were obtained, which are omitted for the lack of space. P1 and P3 are the probability distributions of Y=1 (neutral tone) and Y=3 (normal stress) respectively; P2 is the probability distribution of Y=2 (uncertainty).

Figure 3 plots the simulated 3D and 2D spaces from the regression functions. It illustrates that neutral tone perception is related to pitch in all contexts. As the increase of pitch step and duration step, the perception probability of P1 (neutral tone) increases and the perception probability of P3 (normal stress) decreases.

Numbers in 2D spaces represent the perception probabilities of neutral tone words. The light area on the right side shows that the perception probabilities of neutral tone are higher than those on the left dark area, in which the perception probabilities of normal stress are higher. Two areas are demarcated by a boundary line. As shown in 2D figures, across all three contexts, pitch is significantly related to neutral tone perception.

The absolute value of the slope of the perception boundary is an indication of the extension of correlation. In the on-focus context, the boundary between P1 and P3 is vertical, meaning there is almost no effect of duration and 100% effect of pitch. When combining 2D and 3D results, the effect of duration in isolation is greater than that in the post-focus context. Overall, the effect of duration is much less than that of pitch.

Comparing with those of T1-T0 contrast in [23], the simulated spaces of T2-T0 contrast are quite different, which indicates that both context and tonal space are correlates of perceptual space.

5. DISCUSSION AND CONCLUSION

To study the neutral tone perception, we conducted a perceptual experiment through systematically changing the duration and pitch between normal stressed and neutral tone contrast in three different contexts. The present results confirm that pitch is the most reliable acoustic cue for neutral tone perception, and that duration is a secondary cue. The results also extend the findings in [23] that perception is highly correlated to both context and the tonal spaces of the contrastive tones. Specifically speaking, (i) the amplitude of pitch effect in descending order is isolation > on-focus > post-focus; (ii) across all three contexts, the effect of pitch is greater than that of duration; and (iii) for T1-T0 contrast in [23], the effect of duration is significant in each context. However, in isolation and on-focus contexts, the effect of duration is less than that of pitch. In post-focus context, the effects of duration and pitch are basically the same. For T2-T0 contrast, the effect of duration is significant only in isolation context. But there is no significant effect of duration in on-focus and post-focus contexts. Especially in on-focus context, there is little effect of duration in neutral tone perception. (iv) perceptual patterns or perceptual spaces between the two contrasts of T1-T0 and T2-T0 are both context and tonal space related.

Why is the distribution of perceptual space related to the tonal space between stressed and neutral tones? Comparing Figure 2 in the present study and Figure 3 in [23], the effect of pitch on perception with T1-T0 contrast is less than that with T2-T0 contrast, which implies that a reduced tonal space would cause a reduced contribution of pitch. As a result, the reduced contribution of pitch makes duration comparatively more important. On the contrary, the expanded tonal distance between T2 and T0 would increase the effect of pitch, which consequentially restricts the effect of duration.

Why is the perception dependent on different contexts? It's easy to understand in that the pitch range of a word in neutral tone is affected by context. In focus or isolation context, the pitch range of the target word is larger than that in post-focus position. In Standard Chinese, in post-focus position, the pitch of the whole target word is compressed due to post-focus compression (PFC) [24], leading to a decline of the pitch effect as a result.

The results found in this paper are rather different from Lin's [12], which claimed that the impact of duration is much greater than that of pitch in neutral tone perception. Other researches [10, 11] proposed the greater impact of pitch but didn't investigate the situation at the utterance-level.

Beckman & Edwards [20] and Sluijter & van Heuven [7] all proposed that in non-tonal languages, F_0 is a dependent variable of sentence stress instead of lexical stress; unstressed syllable in word is associated with duration, formants and other features instead of the change of F_0 [21, 22]. Our results stated that in Standard Chinese, as a tone language, F_0 has both functions to express tone and intonation. Therefore, the stress pattern of words, either in isolation as a lexical stress or in utterance as a nuclear pitch accent, is greatly related to F_0, with pitch as the most important correlate. However, when a word is in the post-focus position where the F_0 space is compressed, the impact of pitch is reduced.

A study on neutral tone production is being carried on. We wish to get a whole picture of the acoustic correlates of Standard Chinese from both production and perception. We believe the results will benefit speech understanding and L2 learners.

6. ACKNOWLEDGEMENTS

This work was supported by the National Basic Research Program (973 Program) of China (No. 2013CB329301), CASS innovation project "Key Laboratory of Phonetics and Speech Science" and the joint program between IL, CASS and IL HAS, Hungary "An Interface Study on Focus".

7. REFERENCES

[1] Chao, Y.R. 1922. Gwoyeu Romatzyh or the National Romanization. In Wu, Z. J. (eds), *Linguistic Essays By Yuenren Chao*. Beijing: The Commercial Press, 61-72.

[2] Chao, Y. R. 1979. *Beijing Kouyu Yufa(A Grammar of Spoken Chinese)*. Beijing: The Commercial Press.

[3] Chen, Y. Y., Xu, Y. 2006, Production of Weak Elements in Speech Evidence from F0 Patterns of Neutral Tone in Standard Chinese. *Phonetica* 63, 47-75.

[4] Lu, J. L., Wang, J. L. 2005. Guanyu qingsheng de dingjie (Analysis on the Nature of Neutral Tone). *Contemporary Linguistics* 7(2), 107-112.

[5] Fry, D. B. 1958. Experiments in the perception of stress. *Language and Speech* 1, 126-152.

[6] Sluijter, A. M. C., Heuven, V. J. 1995. Effects of Focus Distribution, Pitch Accent and Lexical Stress on the Temporal Organization of Syllables in Dutch. *Phonetica* 52, 71-89.

[7] Sluijter A. M. C., Heuven, V. J. 1996. Spectral balance as an acoustic correlate of linguistic stress. *Acoustical Society of America* 100 (4), 2471-2485.

[8] Heuven, V. J., de Jonge, M. 2011. Spectral and Temporal Reduction as Stress Cues in Dutch. *Phonetica* 68:120–132.

[9] Sluijter A. M. C., Heuven, V. J. 1997. Spectral balance as a cue in the perception of linguistic stress. *Acoustical Society of America* 101 (1), 503-513.

[10] Cao J. F. 1986. The Acoustic Cues of Neutral Tone syllable in Standard Chinese. *J. Applied Acoustics* 5(4), 1-6.

[11] Wang, Y. J. 2004. The effect of pitch and duration on the perception of the neutral tone in standard Chinese. *ATTA ACUSTICA* 29 (5), 453-461.

[12]Lin, T. 1985. Experiments on Neutral Tone's qualities in Beijing Mandarin. In Lin, T., Wang, L. J. (eds), *Papers on Experimental Phonetics.* Beijing: Peking University Press. 1-26.

[13] Lin, M. C., Yan, J. Z. 1980. Beijinghua qingsheng de shengxue xingzhi. *Dialect* 3: 166–178.

[14] Lin, M. C., Yan, J. Z. 1990. Neutral tone and stress in Mandarin Chinese. *Language Teaching and Linguistic Studies* 3, 88-104.

[15] Lin, T., Wang, W. 1985. Tone perception. *J. Chin. Linguist* 2, 59–69.

[16] Yang, S.A. 1991. The synthetic rules for neutral tone in standard Chinese. *J. Applied Acoustics* 10(1). 12-18.

[17] Lee, W. S. 2003. A Phonetic Study of the Neutral Tone in Beijing Mandarin. *Paper 15th ICPhS*, Barcelona, 1121-1124.

[18] Lin, T., Wang, L. J. 1992. *Course of Phonetics*. Beijing: Peking University Press.

[19] Zhong, X. B., Wang, B., Yang, Y. F., Lv, S. N. 2001. Hanyu yunlvci zhijue yanjiu (The perception analyze on Mandarin Chinese prosodic words). *Xinli Xuebao (Acta Psychologica Sinica)* 33(6), 481-488.

[20] Beckman, M. E., Edwards, J. 1994. Articulatory evidence for differentiating stress categories. *in Keating, Phonological structure and phonetic form. Papers in Laboratory Phonology*, Cambridge University Press, Cambridge. III, 7– 33.

[21] Fry, D. B. The dependence of stress judgments on vowel formant structure. *Proc. 6th ICPhSc* Bethge, 306– 311.

[22] Sluijter, A. M. C., Heuven, V. J., Pacilly, J. J. A. 1997. Spectral balance as a cue in the perception of linguistic stress. *J.acoust. Soc. Am.* 101: 503– 513.

[23] Li, A. J., Gao, J., Jia, Y.,Wang, Y. R. 2014. Pitch and Duration as Cues in Perception of Neutral Tone under Different Contexts in Standard Chinese. *APSIPA* Angkor Wat.

[24] Xu, Y., Chen, S. W., Wang, B. 2012. Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family? *The Linguistic Review* 29: 131-147