Speech Prosody 2016
31 May - 3 June 2016, Boston, USA

The Prosodic Effect of the Neutral Tone to the Preceding Tone

Shanshan Fan[1], *Aijun Li*[1], *Jun Gao*[1], *Ao Chen*[2]

[1] Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China
[2]Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, The Netherlands
fanshan08@163.com, liaj@cass.org.cn, 9894575@qq.com, a.chen@uu.nl

Abstract

Few studies focused on the tonal realization of first syllables in disyllabic words with tonal combination of "canonical tone + neutral tone" in Mandarin. To fill in the void, the present study compared the prosodic characteristics of first syllables between minimal pairs of neutral tone word (NW) and canonical tone word (CW). The results show that (1) For duration, no difference was found. (2) The F0 range is expanded in NW, showing the first syllable in NW is more prominent than that in CW. When we looking at individual tones of the first syllable in NW, the high tone (T1) is raised higher; the falling tone (T4) is expanded larger with higher F0 onset and lower F0 offset; the rising tone (T2) has lower F0 contour, and the dipping tone (T3) is lower, which causes more creaky voices. (3) We find the differences of intensity between CW and NW statistically. However, only the T4T4 shows greater intensity in NW than that in CW. The present study implies that neutral tone, as a weak element, is encoded not only by its own acoustic cues but also by enlarging the preceding F0 range to produce more prominence contrast between the two syllables.

Index Terms: Mandarin, Neutral tone, the preceding tone, production study, prosodic effect

1. Introduction

Mandarin is a typical tone language. There are four lexical tones, which are also named as full tones, including T1 (high-level tone), T2 (low-rising tone), T3 (low-dipping tone) and T4 (high-falling tone). Syllable carrying different lexical tones conveys different meanings, such as /ma1/ 'mother', /ma2/ 'hemp', /ma3/ 'horse', and /ma4/ 'scold'.

Besides full tones, there is a neutral tone (T0), such as the second syllable in /ma1ma0/ 'mother'. The neutral tone is also called the fifth tone, the particular tone sandhi, or the unstressed syllable, which is reduced in duration and F0 contour [5, 22, 16, 25]. Although it is controversial whether there is the lexical stress in Mandarin, it is agreed that the neutral tone is unstressed [10, 11, 12, 17, 24]. A syllable can carry either full tones or the neutral tone, which results in some minimal pairs involving full tones as opposed to neutral tones, such as /tuŋ1ɕi1/ 'east and west' and /tuŋ1ɕi0/ 'things'. The neutral tone never occurs initially nor independently, i.e. the smallest unit for a neutral tone to occur is a disyllabic sequence, and the unstressed position is always non-initial. Therefore, there is always a tone preceding the neutral tone, which determines the F0 contour of the neutral tone (in Figure1).

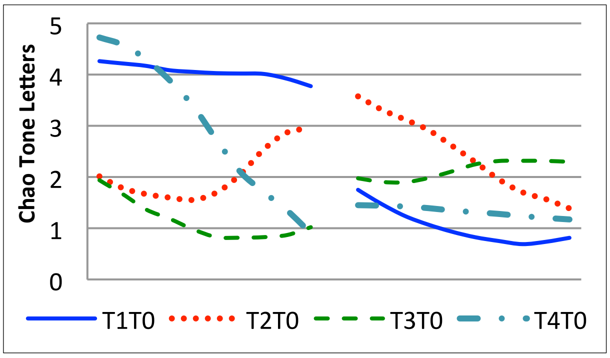

Figure 1, F0 contours of neutral tones after different preceding tones with normalized duration.

Fundamental frequency (F0), duration, spectral tilt, intensity, and vowel reduction are acoustic correlates of neutral tones. In previous studies, both F0 and duration are proved to be more important than other cues for neutral tone perception [19]. When comparing the importance of F0 and duration, F0 is proved to be the most important cue for unstressness [8, 9, 23, 1, 2, 3, 18, 19]. Duration is also important [20, 12, 13, 4], but it is less important than F0 relatively [3, 18]. In continuous speech, the importance of the F0 and duration is highly related to tone combinations, lexical contexts, and focus position [2, 3]. Spectral tilt is also an important cue for unstressed syllables [4, 21], which is less than the function of F0 and duration in Mandarin [21]. Intensity is not reliable for unstressness [14, 24, 8, 4], i.e. the intensity of the syllable with the neutral tone is not necessarily weaker than that with canonical tones [13, 14]. Vowel reduction is reliable in some western Germanic languages [8], but it is not reliable in Mandarin. The using of vowel reductions is highly related to dialect backgrounds or individual habits [12].

Among the studies focusing on the realization of unstressness, most studies compared the phonetic characteristics between the neutral tone and its corresponding full tone. So far, however, it is unclear how the neutral tone may influence the preceding full tone. In our previous studies, applying the duration and pitch features of the neutral tone syllable did not lead to perceptually natural neutral tones. The phonetic cues from the preceding tones might be required when perceiving neutral tones for Mandarin native listeners.

To this end, we recorded disyllabic real words with all possible tone combinations. Then, we analyzed the phonetic differences between the first syllables in neutral tone words (NW) and corresponding canonical tone words (CW). For example, in the neutral tone minimal pair /tuŋ1ɕi1/ 'east and west' and /tuŋ1ɕi0/ 'things', two /tuŋ1/s were compared by

10.21437/SpeechProsody.2016-61

analyzing F0, duration, and intensity. In our assumption, to increase the prominence of the first syllable in NW, the stressed syllable before the neutral tone would carry the larger F0 range, longer duration, and comparatively greater intensity.

2. Methodology

2.1. Recordings

Neutral tone minimal pairs were selected, including all possible tone combinations, except T3T3. In T3T3, tone sandhi changes the underlying T3T3 into T2T3. For example, /lau3tsɿ3/ 'founder of Taoism' is realized as /lau2tsɿ3/ in Mandarin. Therefore, 15 tone combinations were included. For each tone combination, there is one disyllabic real word with neutral tone and its corresponding minimal pair with full tones. The 30 disyllabic words were listed in Table 1.

Table 1, The list of neutral tone minimal pairs.

2nd / 1st	T1/T0	T2/T0	T3/T0	T4/T0
T1	tuŋ1ɕi1/ tuŋ1ɕi0	ia1tʰəu2 ia1tʰəu0	ɕia1tsɿ3/ ɕia1tsɿ0	ɕyŋ1ti4/ ɕyŋ1ti0
T2	mo2ku1/ mo2ku0	ʂɤ2tʰəu2 / ʂɤ2tʰəu0	liɛn2tsɿ3/ liɛn2tsɿ0	fu2tɕʰi4/ fu2tɕʰi0
T3	xuoʂau1/ xuo3ʂau0	ɕiɑŋ3tʰəu2/ ɕiɑŋ3tʰəu0	/	pən3ʂʅ4/ pən3ʂʅ0
T4	piɛn4taŋ1/ piɛn4taŋ0	ʂʅ4tɕʰiŋ2/ ʂʅ4tɕʰiŋ0	li4tsɿ3/ li4tsɿ0	li4tɕʰi4/ li4tɕʰi0

10 Beijing Mandarin native speakers, including 5 males and 5 females, produced all words listed above. Recordings were conducted in a sound attenuating phonetic lab. Each word was read for once in an isolation context. All sounds were collected by using the microphone of AKG C1000S, with the sample rate of 16 kHz.

2.2. Annotation

A professional annotator transcribed all recordings using Praat 5.1 [16]. As shown in Figure 2, there were four tiers, including HZ (Chinese Characters), I-F (initials and finals), PinYin (syllables), and VQ (Voice Quality). The 'sil' represented the silence, and the 'CR' represented the crackly voice.

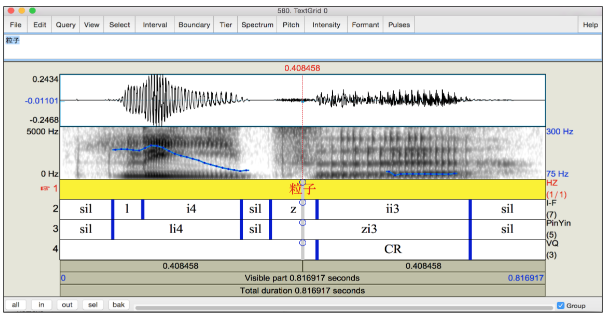

Figure 2: A sample of annotation.

3. Statistical analysis

3.1. Data extraction

Praat 5.1 [16] was used to extract duration, F0, and intensity. Rhymes with creaky voices were excluded from further analyses. When the neutral tone was preceded by a T3, creaky voices occurred more often in NW (87%). This proportion was 20% in CW. T3 was excluded from the statistical analysis due to frequent creaky voices.

For other full tones, 10 equally spaced points along the rhyme were extracted in the unit of Hertz (Hz). F0 values were normalized using semitone (st) for comparison across subjects and genders. The method of normalization was shown in Equation (1).

$$st_{ij} = 12 \times \frac{\lg(f_{ij}/fr_i)}{\lg(2)} \quad (1)$$

In Equation (1), "i" represented speakers' index from 1 to 10. "j" represents the j^{th} F0 value, which was from 1 to 10 as well. And "f_{ij}" represented the specific value of the j^{th} F0 point from the speaker i. "fr_i" represented the referenced frequency of subject i, which is the minimal F0 value of the speaker i.

3.2. Duration

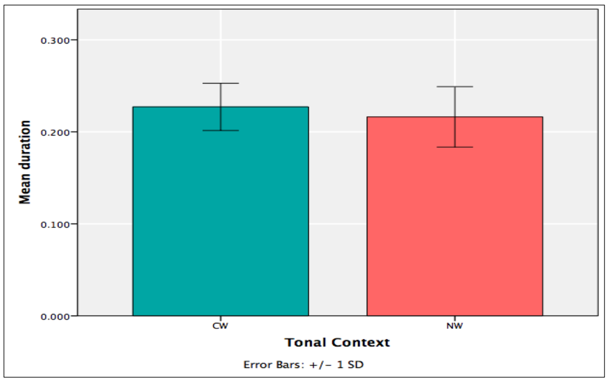

Figure 3: The duration in CW (left) and NW (right) in the unit of second.

Repeated Measures ANOVA was conducted by adding *Tonal Context* as the within-subject factor, including 2 levels: CW, and NW. No effect of *Tonal Context* was found, $F(1, 9) = 1.37$, $p > 0.05$. Although statistically non-significant, first syllables in CW tend to be longer than those in NW on average, which are 0.227 second (sd = 0.047) and 0.215 second (sd = 0.052) correspondently, in Figure 3.

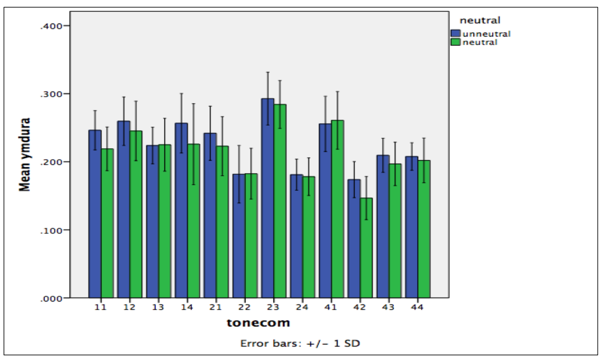

Figure 4: The duration under each tone combination (tonecom) in CW (left bars) and NW (right bars) in the unit of second.

To check the differences of duration under each tome combination, paired samples t-tests were conducted. The duration acted as the dependent variable. Both tone combinations and tonal context were added as group factors. No difference was found under any tone combination, with p > 0.05, in Table 2.

Table 2, The results of paired samples t-tests.

2nd / 1st	T1	T2	T3	T4
T1	t(9)=2.22	t(9) = 1.16	t(9)=0.12	t(9) = 1.9
T2	t(9)=1.45	t(9) = - 0.07	t(8)=0.59	t(8) = 0.21
T4	t(7)= - 0.51	t(9) = 2.22	t(8)= 1.06	t(8) = 0.65

3.3. Fundamental frequencies

3.3.1. F0 range and Pitch level

To analyze data from F0, for each speaker under each tonal context, two parameters were calculated in the unit of semitone, including F0 range, and pitch level. F0 range was calculated by subtracting minimal F0 value from the maximal F0 value. And, the pitch level was calculated by averaging the maximal and minimal F0 values.

Repeated Measure ANOVA were conducted by adding Tonal Context (NW and CW) as the within subject factor. The variables of F0 range, the minimal F0 value, the maximal F0 value, and pitch level were added as the dependent variables one after another. The main effect of F0 range was found, $F(1, 9) = 27.63$, $p < 0.01$, $\eta^2 = 0.76$. In addition, we found the main effect of minimal F0 value, $F(1, 9) = 40.84$, $p < 0.01$, $\eta^2 = 0.82$. We also found the main effect of pitch level, $F(1, 9) = 18.6$, $p < 0.01$, $\eta^2 = 0.67$. But no effect of the maximal F0 value was found, $F(1, 9) = 0.07$, $p > 0.05$. In the pairwise comparisons, first syllables in NW tend to carry larger F0 range ($p < 0.01$), lower pitch level ($p < 0.01$), and smaller minimal F0 value ($p < 0.01$). The F0 range and pitch level in CW and NW were shown in the following Figure 5 and Figure 6 correspondingly.

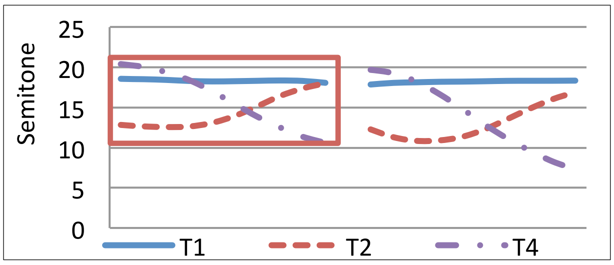

Figure 5:F0 contours of first syllables in CW.

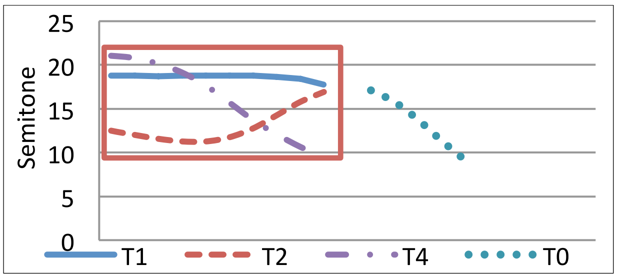

Figure 6:F0 contours of first syllables in NW.

The F0 related parameters between NW and CW, including F0 range, the maximal F0 value (Max F0), the minimal F0 value (Min F0), and pitch level, are listed in Table 3 in the unit of semitone.

Table 3, F0 related parameters between NW and CW.

	F0 range	Max F0	Min F0	Pitch level
NW	11.7	21.1	9.45	14.72
CW	9.8	20.4	10.6	14.98

To check the differences of F0 range between NW and CW under each tone combination, paired samples t-tests were conducted. The results show that the F0 ranges significantly expand in NW by comparing with CW when carrying T1T1 ($t(9)= 3.15$), T1T4 ($t(9) = 5.58$), T2T1 ($t(9) = 3.15$), T4T1 ($t(7) = 7.33$), T4T2 ($t(9) =2.37$), and T4T4 ($t(8) = 2.7$), with $p < 0.05$.

Different patterns were observed when the first syllable carried different tones. The T1 has a higher pitch level when it precedes a neutral tone (the upper-left panel in Figure 7). The F0 range of T4 is expanded with raised F0 peak and lowered F0 valley in NW, comparing with T4 in CW (the upper-right panel in Figure 7). The F0 contour of T2 in NW is underneath that in CW (the bottom panel in Figure 7).

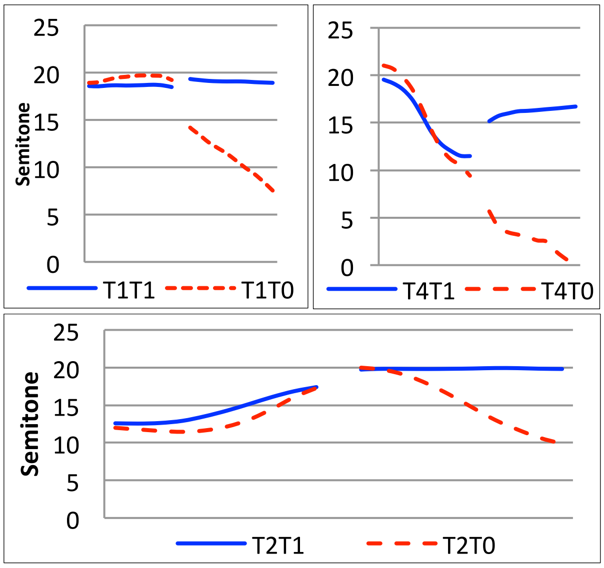

Figure 7: Different patterns of F0 variation in CW (solid lines) and NW (dashed lines).

3.3.2. Similarities

To analyze the variation of the first syllables in NW and in CW, the Euclidean Distances (d) were calculated, as shown in Equation (2).

$$d = \sqrt{\sum_{i=1}^{n} (A_i - B_i)^2} \quad (2)$$

In Equation (2), the "i" represented the number of specific F0 point, from 1 to 10. The "A" referred to the value of the i^{th} F0 point in CW, while the "B" showed the corresponding value in NW, in the unit of semitone. The results were listed in Table 4, where the larger d value means less similarity. With the sequence of decreasing similarity, the ranking is T1, T4, and T2.

Table 4, The Euclidean Distance between first syllables in CW and NW under each tone combination.

2nd / 1st	T1/T0	T2/T0	T3/T0	T4/T0	mean
T1	2.57	0.62	0.75	2.09	1.51
T2	4.44	6.46	7.79	2.72	5.35
T4	4.42	3.09	1.31	2.71	2.88

3.4. Intensity

To check the differences of intensity between CW and NW, Repeated Measures ANOVA was conducted by adding *Tonal Context* as the within-subject factor including CW and NW. We found the significant main effect of *Tonal Context,* $F(1, 9) = 7.83$, $p < 0.05$, $\eta^2 = 0.47$. In pairwise comparison, first syllables in CW have significantly higher intensity than those in NW, $p < 0.05$.

As shown in Figure 8, first syllables in CW are tended to be stronger than those in NW on average, which are 73.97 dB (sd = 2.08) and 73.22 dB (sd = 2.96) correspondently.

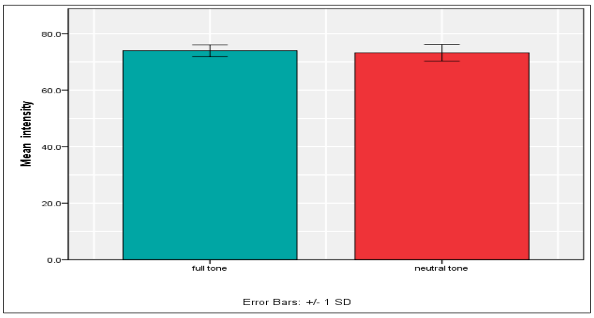

Figure 8, The intensity of first syllables in CW (left) and NW (right) in the unit of dB.

To check the differences of intensity between CW and NW under each tone combination paired samples t-tests were conducted. Differences were not found under all tone combinations, which was only found under the combination of T4T4, $t(8) = 2.84$, $p < 0.05$, as shown in Figure 9.

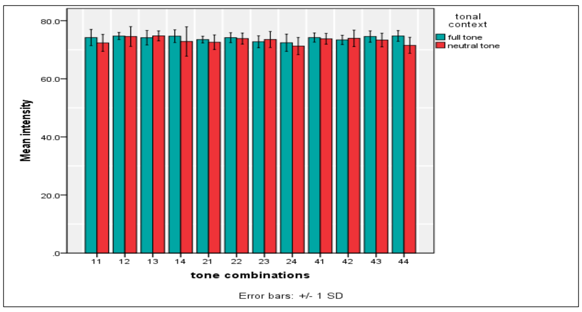

Figure 9: Intensities of first syllables under each tone combination in CW (left bars) and NW (right bars) in the unit of dB.

4. Discussion& Conclusions

The present study compared the phonetic differences of the first syllables in disyllabic words when they are followed by a neutral tone versus a canonical tone.

The results show that there is no significant difference between first syllables in NW and CW. We find the effect of intensity, but it is only found under the tone combination of T4T4 when we considered tone combinations. The F0 range of the first syllable is expanded in NW compared with that in CW. The results disclosed that the neutral tone in Mandarin is encoded not only by varying the phonetic cues of its own, but also by expanding the F0 range of the proceeding tone to introduce more prominent contrast between the two syllables.

However, the influences of neutral tones on the first syllables are highly tone specific. Specifically, when the first syllable carries high-level tone (T1), the F0 contour is raised. When the low-rising tone (T2) is the preceding one, the F0 contours are lowered, compared with that in CW. When it carries high-falling tone (T4), F0 maximum is raised and minimum is lowered. Whether the differences between first syllables are from the effects of coarticulation or not needs further research.

Besides the F0 range, the maximal value, the minimal F0 value, the pitch level, the Euclidean distance between first syllables in NW and in CW were calculated. The significant differences of the minimal F0 value and the pitch level were found. But no significant difference was found for maximal value. Therefore, the main way to expand the F0 range is the decrease of the minimal F0 value instead of increasing the peak value. The lower pitch level might be another way to introduce the following syllable with neutral tone. This interprets the frequent creaky voices in NW with T3 (low-dipping tone) as the preceding tone. The declination of the F0 valley and pitch level of T3, which is already located at the bottom of the pitch, causes the high frequently creaky voice than other preceding tones.

The results of the Euclidean distance show that: 1) when the first syllable carries T2, the distance between first syllables in NW and CW is the largest; 2) when the first syllable carries T1, the distance between first syllables is the smallest; 3) when the first syllable carries the high-falling tone (T4), the distance is intermediate.

Our coming working is to look into the situation in utterances. We are checking the prosodic cues of NW and CW when they are embedded in sentences with different information structures.

5. Acknowledgements

This work was supported by the National Basic Research Program (973 Program) of China (No. 2013CB329301), the CASS innovation project “Key Laboratory of Phonetics and Speech Science”, and the China Exchange Program (CEP) between KNAW and CASS (No. 11CDP004).

6. References

[1] A. Li, *Report in CASS Innovation Program Annual Summing up Meeting*. Beijing, Nov. 2013.

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

[3] A. Li, and S. Fan, "Correlates of Chinese Neutral Tone Perception in Different Contexts," Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015), Glasgow, Scotland, 2015.

[4] A. M. C. Sluijter, and V. J. van Heuven, "Spectral balance as an acoustic correlate of linguistic stress," Journal of the Acoustic of America, 100, pp. 2471-2485, 1996.

[5] C. Luo, and J. Wang, *An Outline of General Phonetics*. Beijing: The Commercial Press, 2000.

[6] C. N. Li, and S. A. Thompson, "The acquisition of tone in Mandarin-speaking children," *Journal of Child Language*,vol. 4, no. 2, pp. 185–199, 1976.

[7] C. Li, "Remarks on the Underlying Representations of Neutral-Tone Words in Mandarin Chinese," *Stud. English Lit. Linguist*, pp. 189–199, 1987.

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

[9] J. Cao, "The Analysis of Neutral Tone's Qualities in Mandarin", *Applied Acoustics*, vol. 5, no. 4, pp. 16, 1986.

[10] J. Lu, and J. Wang, "Analysis on the Nature of Neutral Tone", *Contemporary Linguistics*, vol. 7, no. 2, pp. 107-112, 2005.

[11] J. Wang, "The representation of the neutral tone in Chinese Putonghua," *In J. Wang and N. S. H. Smith, "Studies in Chinese Phonology,"* Berlin/New York: Mouton de Gruyter, pp. 157-184, 1997.

[12] M. Lin, J. Yan, and G. Sun, "The experiment on disyllabic words with normal stress in Beijing Mandarin", *Dialect*, 1, pp. 57-73, 1984.

[13] M. Lin, J. Yan, "Neutral tone and stress in Mandarin Chinese", *Lang Teaching &Linguis Studies*, 3, pp. 88-104, 1990.

[14] M. Lin, *Experimental Research on Chinese Intonation*. Beijing: China Social Sciences Press, 2012.

[15] M. Zhou, *Phonology Analysis of Vowel Reduction of Standard Chinese,*Master thesis, Beijing Forestry University, Beijing.

[16] P. Boersma, and D. Weenink, "Praat: doing phonetics by computer (Version 5.1.25) [Computer program]", Retrieved January 20, 2010. From http://www.praat.org/

[17] S. Duanmu, *The Phonology of Standard Chinese*. New York: Oxford University Press, 2000.

[18] S. Fan, A. Li, J. Gao, and A. Chen, "A Perceptual Study on Chinese Neutral Tone and Its Acoustic Correlates", *in the 17th O-COCOSDA*, September, Phuket, 2014.

[19] S. Fan, A. Chen, and A. Li, "The Recognition of Neutral Tone Across Acoustic Cues", *in the 18th OCOCOSDA*, Oct. , Shanghai, 2015.

[20] T. Lin, "Experiments on Neutral Tone's qualities in Beijing Mandarin", in T. Lin, L. Wang.,*Working papers on experimental phonetics*. Peking University Press, 1985.

[21] X. Zhong, B. Wang, Y. Yang, S. Lv, "The perception analyze on Mandarin Chinese prosodic words", *ActaPsychologicaSinica*, vol. 33, no. 6, pp. 481-488, 2001.

[22] Y. R. Chao,*A Grammar of Spoken Chinese.* Beijing: The Commercial Press, 1979.

[23] Y. Wang, "The function of pitch and duration in Mandarin Chinese neutral tone perception", *Chinese Journal of Acoustics*, vol, 29, no. 5, pp. 453-461, 2004.

[24] Z. Wu, and M. Lin, *A Course in Experimental Phonetics*. Beijing: Higher Education Press, 1989.

[25] Z. Wu, "Introduction to Contemporary Phonetics", Beijing: Sinolingua, 1992.