# Prosodic grouping in Chinese trisyllabic structures by multiple cues – tone coarticulation, tone sandhi and consonant lenition

*Wei Lai, Jianjing Kuang*

Department of Linguistics, University of Pennsylvania
<weilai, kuangj>@sas.upenn.edu

## Abstract

Traditionally, the prosodic domain as has been called 'foot' in Mandarin Chinese is considered to be derivable from the application of Tone 3 sandhi rule. This study investigated the internal prosodic grouping of Chinese trisyllabic structures by examining multiple cues in parallel – tone coarticulation, tone sandhi application and consonant lenition. Analyses by tone coarticulation and consonant lenition were consistent with each other, both showing a grouping effect between the former two syllables in a trisyllabic structure. This pattern is especially evident on the fast speech rate condition. However, these analyses contradicted the analysis by tone sandhi, in that tone sandhi application indicated a prior grouping effect between the latter two syllables in trisyllabic nominal phrases and verbal phrases. The finding that tone sandhi domain violated the minor rhythmic unit reflected by consonant lenition and tone coarticulation suggested that foot formation and tone sandhi application might not be the same process in Mandarin. It was argued that "foot" was encoded and reflected by rhythmically organized phonetic cues such as pitch and timing, not by tone sandhi.

**Index Terms**: tone sandhi, foot, prosody, Mandarin Chinese, tone coarticulation, consonant lenition

## 1. Introduction

In metrical phonology, metrical elements (moras, syllables) are grouped into constituents, or 'feet'. It is normally agreed that a foot in Mandarin usually consists of two or more syllables, and the disyllabic ones are prevalent [1, 2]. For disyllabic words or phrases, the formation of a foot can be easily derived from their lexical forms.

However, foot formation for structures with an odd number of syllables is problematic. A foot containing only one syllable in Mandarin is called a 'degenerate foot', an d a degenerate foot is generally disfavored because it is too light. One solution is to merge it with a neighboring binary foot to form a superfoot [3]. It is sometimes difficult to identify the internal prosodic structure of a superfoot, due to the absence of salient cues to mark foot boundaries.

One well elaborated criterion for foot identification in Mandarin is to find out the prosodic domain where tone sandhi applies, following the tradition of [4-6]. Mandarin has four tones: T1 (level /55/), T2 (rising /35/), T3 (low-rise /214/) and T4 (falling /51/). Tone sandhi in Mandarin refers to a phonological process that changes the first tone of two consecutive low tones (T3) into a rising tone (T2). Based on richly documented empirical observation of tone sandhi application, [5] puts forward the tone-sandhi based *Foot Formation Rule* in Mandarin, as cited in the following:

*Foot Formation Rule*

1. *Link immediate constituents into disyllabic feet;*
2. *Scanning from left to right, string together unpaired syllables into binary feet;*
3. *Join any leftover monosyllable to a neighboring binary foot to form a 'superfoot' according to the direction of syntactic branching.*

This claim indicates two important assumptions: foot formation is determined by both the binary-foot rule and syntactic branching; and tone sandhi application shares the process of foot formation. Pushing these assumptions even further, [7] claims that the prosodic unit "foot" in Mandarin is purely a theoretical construct that has no other purpose than to define the scope of the tone sandhi domain.

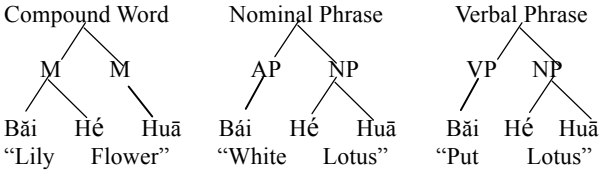
Meanwhile, on the phonetic side, "foot" is considered to be rhythmically encoded by acoustic cues such as pitch and timing. Since prosodic grouping can be affected by prominence, one line of research on the acoustic realization of minor rhythmic units is to study the acoustic correlates of "stress" in Mandarin [8-13]. Besides, there is another line of research that explores the phonetic cues of prosodic grouping in minor rhythmic units without the mediation of prominence. Among these cues, consonant lenition has been reported as an indicator of minor rhythmic units, as heavier lenition often occurs at smaller prosodic boundaries. An empirical observation is that in Mandarin disyllabic structures, onset consonants wrapped within the structure are much less "consonantal" than the structure-initial ones: they are shorter in duration, and are more likely to become voiced if voiceless [14]. Results from corpus studies also reveal that the duration of onset consonants becomes shorter as the prosodic boundary they are located at becomes smaller. This effect is especially clear for fricatives and affricates [15].

Another phenomenon in Mandarin that frequently occurs to minor rhythmic units is tone coarticulation. [16] finds that tonal contours can be largely misshaped within disyllabic or trisyllabic structures, especially in "conflicting" tonal context. [17] finds that a rising tone embedded between two high tones in a trisyllabic compound word can be greatly reduced due to its word-mediate (prosodically weak) position. These findings indicate that similar to heavier consonant lenition, heavier tone coarticulation is also an indicator for smaller prosodic domains.

## 2. Background

This study aims to make use of the above mentioned cues to explore one of the most disputable issues in the research of Chinese phonology – the foot formation or prosodic grouping within trisyllabic structures of different syntactic categories. Since trisyllabic structures are the minimal structures with odd-numbered syllables that can have different syntactic junctures and branching, the internal prosodic grouping of such structures is a natural entry to solving phonology-syntax mapping issues. In Mandarin there are mainly three types of trisyllabic structures: left-branching compound words (CW), right-branching nominal phrase (NP) and right-branching

verbal phrase (VP), as illustrated by the following examples. Right-branching compound words and left-branching phrases also exist, but they are largely disadvantaged in terms of quantity and productivity compared to their prevalent counterparts [1, 2], thus they are not discussed in this study.




In general, there are four different opinions on the prosodic grouping of these three structures: a) CW and NP grouped as [2+1] while VP is grouped as [1+2] [18]; b) CW is grouped as [2+1] while NP and VP are [1+2] [1]; c) CW is [2+1], VP is [1+2] and NP is somewhere in between [19]; d) the three structures are all the same in terms of being a superfoot [3]. One cause for such a disputation is that these studies adopted different criteria to evaluate prosodic grouping, which includes syntactic branching [1], Tone 3 sandhi application [3, 19], reference to tone sandhi in another dialect [18], pitch range, pause and lengthening [19]. The main opinions and criteria used in different studies to address this issue are summarized in Table 1.

Table 1. Previous research on the prosodic grouping of Chinese trisyllabic structures

| Citation | CW | NP | VP | Criteria |
|---|---|---|---|---|
| [18] | 2+1 | | 1+2 | TS in shanghai dialect |
| [1] | 2+1 | 1+2 | | Syntactic branching |
| [19] | 2+1 | In between | 1+2 | TS in a T3+T3+T3 sequence, pitch range, pause, lengthening |
| [3] | The same (a superfoot) | | | TS in a T3+T3 sequence embedded in a trisyllabic structure |

Even analyses based on Tone 3 sandhi can lead to different conclusions. [3] considers the three structures to be prosodically identical in that tone sandhi applies without failure for any T3+T3 sequence embedded in a trisyllabic structure, no matter the extra non-T3 syllable is added to the right or to the left. However, this analysis is problematic for a T3+T3+T3 sequence [19] whose tone sandhi application is structure-dependent: a left-branching structure would surface as T2+T2+T3, while in a right-branching structure would surface as either T2+T2+T3 or T3+T2+T3 (see section 4.1).

Hereafter, we set off to examine phonological cues (tone sandhi) as well as phonetic cues (consonant lenition and tone coarticulation) in Chinese trisyllabic structures to get an integrated understanding of their internal prosodic grouping.

# 3. Experiment I: Tone coarticulation

## 3.1. Hypothesis

According to [13], tone coarticulation is found heavy within small prosodic unit such as disyllabic or trisyllabic words, especially when the tone combination is “conflicting”. Our first experiment is to investigate trisyllabic structures of “conflicting” tone sequences (T2+T2+T2 and T4+T4+T4), and compare the tone coarticulation between the former pair of syllables and that between the latter pair of syllables. By this analysis, whichever pair of syllables that undergoes heavier coarticulation is supposed to be grouped more tightly.

## 3.2. Materials

Trisyllabic structures with tonal sequences of T2+T2+T2 and T4+T4+T4 were constructed respectively for the three syntactic categories (CW, NP, VP). Other structures of different tone combinations and syllable counts were used as fillers to separate the target structures.

7 native speakers of Mandarin Chinese were recruited to read the wordlist. The speakers were asked to read each word 6 times, with the first 3 times in a slow speech rate and the second 3 times in fast speech rate. In all, we have 7 (speakers) * 3 (repetition) * 2 (speech rates) * 3 (syntax) * 2 (tones) = 252 tokens for the test of tone coarticulation.

The speakers were recorded in a sound-treated booth in the Phonetics Lab at University of Pennsylvania. They were seated with a SHURE WH320 Condenser microphone positioned approximately 10 cm from her mouth. The recordings were made with Audacity using a sampling rate of 32 bit /44.1 kHz.

## 3.3. Annotation

We manually annotated the tonal nuclei by defining the feature points of the tone, as shown in Fig. 1.

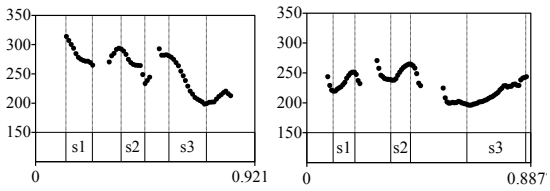



*Fig. 1. Annotation of tone nuclei for T4+T4+T4 (left) and T2+T2+T2 (right) tonal sequences (x-axis: Hz; y-axis: s)*

The tonal nuclei were analyzed using Xu’s script [20]. 10 $F_0$ points on the contour were evenly extracted from each tone nucleus. Within speaker normalization was done for $F_0$ values using *z-score* by equation (1), where $x$ stands for specific $F_0$ values, $\mu$ and $\sigma$ respectively stand for the mean and deviation of all $F_0$ values from a certain participant.

$$z=(x-\mu)/\sigma \quad (1)$$

We then calculated the pitch reset $\Delta p_i$ between each pair of consecutive tones as an indicator of the coarticulation effect. To make $\Delta p_i$ more comparable between tones, we did the subtraction in opposite directions for $\Delta p$ of T2 sequences and $\Delta p$ of T4 sequences:

$$\text{For T2+T2+T2 sequence, } \Delta p_{i\,(i=1,\,2)} = p_{Ei} - p_{B(i+1)} \quad (2)$$

$$\text{For T4+T4+T4 sequence, } \Delta p_{i\,(i=1,\,2)} = p_{B(i+1)} - p_{Ei} \quad (3)$$

In (2) and (3), $p_{Ei}$ stands for the pitch value at the end of a preceding tone; $p_{B(i+1)}$ stands for the pitch value in the beginning of a following tone. $\Delta p_1$ and $\Delta p_2$ respectively stands for the pitch reset between the former two syllables and the latter two syllables. Note that this step is not equivalent to taking the absolute value of the difference, because both positive and negative differences are observed for each tonal sequence.

## 3.4. Results

Fig. 2 shows the pitch contours averaged across 7 speakers for structures of 3 syntactic categories (CW, NP, VP), with

each category containing 2 tonal sequences (T2+T2+T2, T4+T4+T4). We can observe that pitch contours for the first two tones are connected and continuous, indicating much assimilation and carry-over effect between these two tones. By contrast, the latter two tones are less coarticulated, reflected by a large pitch reset in between. This contrast is especially evident for T2+T2+T2 sequences on both speech rate conditions and T4+T4+T4 sequences on the fast speech rate condition. The syntactic category has small influence mostly on $F_0$ range of the first and second tones, but it does not seem to change the coarticulation pattern for any of the particular structures.

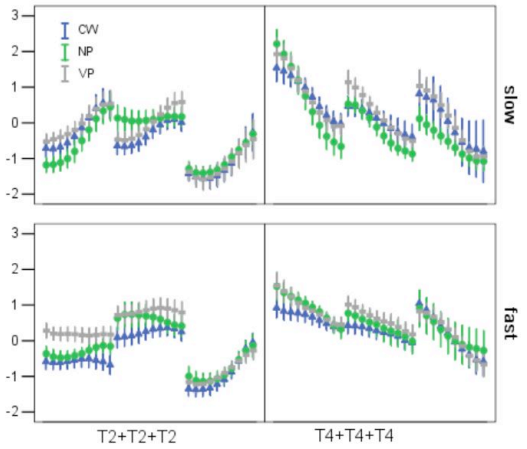



*Fig. 2. Tone contours of T2+T2+T2 and T4+T4+T4 structures of different syntactic categories*

Table 2 shows the values of pitch reset between the former two tones and that between the latter two tones on all conditions of syntactic categories and speech rates. We can see that the coarticulation effect is heavier between the former two syllables than between the latter two syllables, reflected by smaller values of pitch reset. The difference between $\Delta p_1$ and $\Delta p_2$ is obviously larger on the fast speech rate condition, suggesting a clearer grouping pattern of [2+1].

Table 2. Pitch reset between the former two syllables ($\Delta p_1$) and the latter two syllables ($\Delta p_2$) and results by t-test

| Speech Rate | Syntax | $\Delta p_1$ | $\Delta p_2$ | t-test |
|---|---|---|---|---|
| Slow | CW | 0.82 | 1.32 | sig < .01 |
| | NP | 0.75 | 1.22 | sig < .05 |
| | VP | 1.12 | 1.75 | sig < .01 |
| Fast | CW | -0.34 | 1.35 | sig < .001 |
| | NP | -0.17 | 1.16 | sig < .001 |
| | VP | 0.01 | 1.30 | sig < .001 |

Paired sample t-test reveals that the difference between $\triangle p_1$ and $\triangle p_2$ is significant for structures of all syntactic categories in both slow and fast speech rates. Syntax has an influence on the separate value of $\triangle p_1$ and $\triangle p_2$, but it does not change the pattern $\triangle p_1 < \triangle p_2$. In all, the analysis by tone coarticulation reveals a general prosodic pattern of [2+1] for trisyllabic structures, along with a subtle and gradient effect from syntactic categories when it is on the slow speech rate condition.

# 4. Experiment II: Tone Sandhi

## 4.1. Hypothesis

According to [5], the Tone 3 sandhi rule applies cyclically following the branching of syntax. Therefore, an underlying T3+T3+T3 sequence can generate two different surface forms as the syntactic branching of the structure differs, which is demonstrated in the following:

**[2+1]:** [[T3+T3]+T3] [T2+T3+T3] T2+T2+T3

**[1+2]:** [T3+[T3+T3]] [T3+T2+T3] T3+T2+T3

Note that in the reality the relationship between syntactic branching and surface tone combination is not one-to-one mapping. Tone sandhi applies obligatorily to the former two syllables in left branching structures (CW), resulting in a surface pattern of T2+T2+T3, while tone sandhi is optional for the second syllable in right branching structures (NP, VP), so both T2+T2+T3 and T3+T2+T3 are acceptable.

Our second experiment is to record trisyllabic NP and VP structures that have an underlying T3+T3+T3 tone sequence, and check the distribution of the two acceptable surface patterns: T2+T2+T3 and T3+T2+T3. By this analysis, more T2+T2+T3 indicates a prosodic pattern of [2+1], while more T3+T2+T3 indicates a prosodic pattern of [1+2].

## 4.2. Materials

We revised two old examples ("zhǐlǎohǔ" and "ruǎnzǐcǎo") from [5] to form two groups of T3+T3+T3 sequences. Each group contains two structures of different syntactic categories: NP and VP. Filler structures of different tone combinations and syllable counts were inserted between target structures.

10 native Mandarin speakers were recruited to read the wordlist in both slow and fast speech rates, with 3 times of repetition in each speech rate. In all, there are 10 (speakers) * 3 (repetition) * 2 (rates) * 3 (syntax) * 2 (groups) = 360 tokens for the test of tone sandhi.

The setting of recording was the same as described in 3.2. For each token, the surface tone pattern was categorically identified and counted by the first author.

## 4.3. Results

Table 3 shows the number of tokens for each generated surface pattern on different conditions of syntactic categories and speech rates in each of the 6 repetitions. Since tone sandhi application for CW is obligatorily T2+T2+T3 by phonological rules, we focus mainly the distribution of surface patterns for NP and VP structures.

Table 3. Distribution of surface patterns generated by tone sandhi application for underlying T3+T3+T3 structures

| Speech Rate | | Slow | | | Fast | | |
|---|---|---|---|---|---|---|---|
| Repetition | | 1 | 2 | 3 | 1 | 2 | 3 |
| NP | T2+T2+T3 | 1 | 1 | 1 | 1 | 1 | 1 |
| | T3+T2+T3 | **19** | **19** | **19** | **19** | **19** | **19** |
| VP | T2+T2+T3 | 2 | 1 | 1 | 1 | 1 | 1 |
| | T3+T2+T3 | **18** | **19** | **19** | **19** | **19** | **19** |
| CW | T2+T2+T3 | By phonological rule | | | | | |

Table 3 shows that for NP and VP structures, T3+T2+T3 is overwhelmingly preferred than T2+T2+T3 in every repetition,

in both slow and fast speech rates. This result indicates that by the analysis of tone sandhi, trisyllabic VP and NP structures tend to foot the latter two syllables together, while trisyllabic CW structures tend to foot the former two syllables together. This contradicts our previous analysis by tone coarticulation in section 3.

## 5. Experiment III: Consonant Lenition

### 5.1. Hypothesis

According to [15], consonants located at the edge of smaller prosodic junctures will undergo heavier lenition than those located at the edge of larger prosodic junctures. The lenition can be reflected by both shorter duration and a change in spectrum quality. The third experiment is to compare the lenition of the second onset consonant with the lenition of the third onset consonant by duration, to determine whether the former two syllables or the latter two syllables are grouped more tightly. By this analysis, shorter consonant duration between the former two syllables would support a prosodic pattern of [2+1], while shorter consonant duration between the latter two syllables would support a pattern of [1+2].

### 5.2. Materials

We constructed 3 groups of trisyllabic structures, with each group containing 3 structures of different syntactic categories (CW, NP, VP) but identical segments. For example, as shown in section 2, the same syllable sequence "*baihehua*" can cover all the 3 syntactic categories when combined with different tones. To facilitate the observation of consonant lenition, the onset of the last two syllables are constructed to be fricatives or affricates. Structures of other syllable counts were inserted as fillers to separate the target structures.

12 native Mandarin speakers aged from 23 to 40 were recruited for this experiment. Speaker were asked to read each word for 6 times, with the first 3 times in slow speech rate and the last 3 times in fast speech rate. In all, we have 12 (speakers) * 3 (repetition) * 2 (rates) * 3 (syntax) * 3 (group) = 648 tokens. The setting of recording is the same as described in 3.2. For each token, we extracted the duration of the second and third onset consonants.

### 5.3. Results

Fig. 3 shows the duration of the second onset consonant (C2) and third onset consonants (C3) for structures of all three syntactic categories.

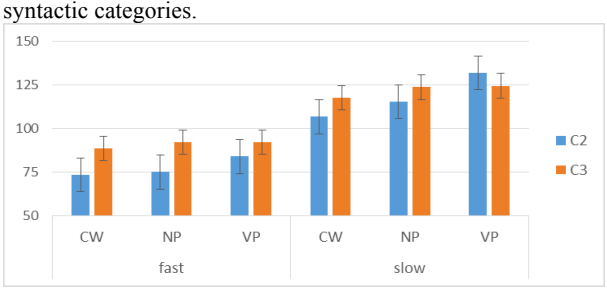



*Fig. 3. Duration of the second and third consonant in CW, NP and VP trisyllabic structures (y-axis: ms)*

We can see that on the slow speech rate condition, syntax is laying a gradient effect on the three types of structures. For CW and NP structures, C2 is shorter than C3, indicating tighter grouping effect between the former two syllables; while for VP structures, C3 is shorter than C2, indicating tighter grouping effect between the latter two syllables. Meanwhile, for CW and NP that both show a prosodic pattern of [2+1], we can also observe a gradient difference on the slow speech rate condition: the first two syllables are grouped tighter for CW and looser for NP. However, on the fast speech rate condition, all the three structures clearly show shorter duration for C2 than C3, regardless of their syntactic categories. This indicates that on the fast speech rate condition, the syntactic effect is minimized and the default prosodic pattern for trisyllabic structures becomes dominant, which is [2+1].

Paired sample t-test shows that the difference of duration between C2 and C3 is significant for all three structures on the fast speech rate condition and two out of three structures (CW, NP) on the slow speech rate condition (sig.<.05), while no significant difference in duration is detected between C2 and C3 for VP on the slow speech rate condition (sig.>.05). This result indicates that the default prosodic pattern for trisyllabic structures in Mandarin is [2+1], but on the slow speech rate condition, this grouping pattern might be influenced by syntactic junctures in a gradient way. This analysis is consistent with the analysis by tone coarticulation, but inconsistent with the analysis by tone sandhi.

## 6. Discussion and Conclusion

The present study investigates the prosodic grouping of Chinese trisyllabic words by examining multiple cues – tone coarticulation, tone sandhi and consonant lenition. We find that the analysis by tone coarticulation is consistent with the analysis by consonant lenition, both suggesting a default prosodic pattern of [2+1] for trisyllabic structures of different syntactic categories in Mandarin. This prosodic pattern can be affected by syntactic junctures in a gradient way on the slow speech rate condition, turning the prosodic pattern of VP structures closer to [1+2]. These analyses reliably suggest that phonetic cues such as coarticulation and duration are sensitive indicators of prosodic grouping and rhythmic organization even for relative small prosodic units.

However, the analysis by tone sandhi indicates a different pattern of prosodic grouping of [1+2] for trisyllabic NP and VP structures in both slow and fast speech rates, which is not consistent with the analyses by tone coarticulation and consonant lenition. The finding that the tone sandhi domain violates the prosodic domain of "foot" defined by pitch (tone coarticulation) and timing (consonant lenition) indicates that tone sandhi application might be a different prosodic mechanism from foot formation. As earlier elaborated in [21], tone sandhi application sometimes can cross distant prosodic boundaries signaled by lengthening and pausing. Therefore, it can be problematic to use tone sandhi as the criterion to define a minimal rhythm unit like a foot.

## 7. Acknowledgement

This study is supported by UPenn faculty research fund to Professor Jianjing Kuang.

## Reference

[1]. Feng, S. (1998). 论汉语的"自然音步" [On the "natural foot" of Chinese Mandarin]. *《中国语文》[Zhongguo Yuwen: Journal of Chinese Linguistics]*. No. 1: 40-47. (Beijing, China).

[2]. Duanmu, S. (1999). 重音理论和汉语的词长选择. [Stress theory and the preference for word length in Chinese Mandarin]. *《中国语文》*[*Zhongguo Yuwen: Journal of Chinese Linguistics]*. No. 4: 246-254. (Beijing, China).
[3]. Shih, C. (1997). Mandarin third tone sandhi and prosodic structure. Linguistic Models, 20, 81-124.
[4]. Chen, M. Y. (1991). An overview of tone sandhi phenomena across Chinese dialects. Journal of Chinese Linguistics Monograph Series, 111-156.
[5]. Shih, C. L. (1986). The prosodic domain of tone sandhi in Chinese (Doctoral dissertation, University of California, San Diego).
[6]. Hung, T. T. (1989). Syntactic and semantic aspects of Chinese tone sandhi. Indiana University Linguistics Club Publications.
[7]. Chen, Matthew Y. "What must phonology know about syntax." The phonology-syntax connection (1990): 19-46.
[8]. Wang Y. J., "The Perception of Disyllabic Word Stress of Chinese Speech in Utterance", Acta Acustica, 6: 534-539, 2003.
[9]. Zhong X. B., "Perception of Sentence Prominence in Chinese and Its Acoustic Parameter", Ph.D thesis, 2000.
[10]. Shen J., "The Principle of Chinese Accent (simple version)", Linguistics Researches, 8: 10-15, 1994.
[11]. Wang B., "The Pitch Movement of Stressed Syllable in Chinese Sentences", Acta Acustica, 3: 234-240, 2002
[12]. Lai, C., Sui, Y., & Yuan, J. (2010). A corpus study of the prosody of polysyllabic words in Mandarin Chinese. *Paper presentedat Speech Prosody*.
[13]. Lin M. C., Yan J. Z. and Sun G. H., "Preliminary Experiment on Normal Stress of Disyllables in Beijing Mandarin", Dialect, 1: 57-73, 1984.
[14]. Xu, Y. (1986). 普通话音联的声学语音学特性 [Acoustic-phonetic characteristics of juncture in Mandarin Chinese]. *《中国语文》[Zhongguo Yuwen: Journal of Chinese Linguistics]* No. 4: 353- 360 (Beijing, China)
[15]. Yuan, Jiahong, and Mark Liberman. "Investigating Consonant Reduction in Mandarin Chinese with Improved Forced Alignment." Sixteenth Annual Conference of the International Speech Communication Association. 2015.
[16]. Xu Y. Production and perception of coarticulated tones[J]. The Journal of the Acoustical Society of America, 1994, 95(4): 2240-2253.
[17]. Shih, C. (2005). Understanding phonology by phonetic implementation. INTERSPEECH. 2005.
[18]. Duanmu, S. (2004). 汉语的节奏 [The rhythm of Chinese Mandarin]. *《当代语言学》 [Dangdai Yuyanxue: Journal of Modern Linguistics]* No. 4: 203-209 (Beijing, China).
[19]. Wang, H. (2000). 汉语的韵律词与韵律短语. [The prosodic word and prosodic phrase of Chinese Mandarin]. *《中国语文》[Zhongguo Yuwen: Journal of Chinese Linguistics]*. No. 6: 525-536 (Beijing, China).
[20]. Xu, Y. (2005). TimeNormalizeF0.praat.
[21]. Kuang, J & Wang, H. (2006). 连上变调在不同韵律层级上的声学表现 [T3 sandhi at the boundaries of different prosodic hierarchies]. 《中国语音学报》 [Zhongguo Yuyin Xuebao: Journal of Chinese phonetics], Volume 1,