

# How does phonetics interact with phonology during tone sandhi?

Bijun Ling

Tongji University  
lingbijun@icloud.com

## ABSTRACT

This paper investigated the phonetics and phonology of consonant–f<sub>0</sub> interaction in Shanghai Wu. Bi-syllabic compound nouns, which form tone sandhi domain, were elicited within template sentences with two factors controlled: lexical tones (T1[HM], T3[LM], T5[LMq]) and consonant types (obstruents & nasals). Results showed that although the base tone contrast of the second syllable is neutralized by phonological tone sandhi rules, the onset f<sub>0</sub> of the second syllable with low tones (T3) is significantly lower than that with high tone (T1). Furthermore, such difference cannot be just attributed to the consonant perturbation, because it also exists when the consonant (i.e. /m/) is the same for all three tones. Therefore, we propose that speakers voluntarily control their articulatory settings, so as to maintain the laryngeal contrast of the second syllable. The observed f<sub>0</sub> pattern is the product of the interaction of phonological rules and phonetic mechanism.

**Keywords:** f<sub>0</sub> perturbation, tone sandhi, Shanghai Wu

## 1. INTRODUCTION

It is well-known that obstruents affect the fundamental frequency (f<sub>0</sub>) of the following vowel. In non-tonal languages, f<sub>0</sub> is higher following a voiceless obstruent than following a voiced one [7][17]. In tonal languages where f<sub>0</sub> differences indicate lexical contrasts, such an influence is often manifested as co-occurrence restrictions between tone and the onset consonant of tone-bearing syllables [4]. Normally, voiced obstruents tend to occur with a low tone while their voiceless counterparts occur with a high tone [8-9]. Therefore, the interaction of f<sub>0</sub> and obstruent onset is governed by the phonological rules in tonal languages while it is controlled by phonetic mechanism in non-tonal languages. Over the past decades, many studies have been done to investigate the mechanisms of consonant articulation that account for the local f<sub>0</sub> perturbation, i.e. the phonetic mechanism [10-12][16][18-19] or the extent to which consonant–f<sub>0</sub> interaction has been exaggerated or restrained beyond physiological necessity, i.e. the phonological effect [13-15]. However, there has been little study on the possible link between the two levels:

whether the phonological system affects the phonetic interaction of consonant and f<sub>0</sub>.

Shanghai Wu, a northern Wu dialect of Chinese, offers a good study case for this research question. Shanghai Wu has five lexical tones, which can be described by three features [27]: F<sub>0</sub> contour: falling (T1) and rising (T2-T5); Tonal register: high (T1, T2, T4) and low (T3, T5); and Duration: long (T1-T3) and short (T4, T5). They exhibit interesting co-occurrence patterns with both the onset and coda of the tone-bearing syllable. Syllables with voiceless onsets only allow tones that start in the high register, i.e. T1, T2 and T4; while voiced onsets co-occur with tones that start in the low register, i.e. T3 and T5. Interestingly, the sonorant consonants could occur with both high register tones and low register tones. Furthermore, the short tones (i.e. checked tones) only co-occur with glottal-closed syllables, while the long tones (i.e. smooth tones) co-occur with open syllables or nasal-closed syllables. The co-occurrence restrictions on tonal categories with consonants and syllable structures are illustrated in Table 1.

**Table 1:** Co-occurrence restrictions on tonal categories with consonants and syllable structure. The number in the square brackets is the value of citation tones in Shanghai Wu (using Chao’s five-level numerical scale, which divides a speaker’s pitch range into five scales with 5 indicating the highest and 1 the lowest).

Consonant type	Obstruent		Sonorant (nasal & labial)
	voiceless	voiced	
CV(ŋ)	T1[53], T2[34]	T3[13]	T1[53], T2[34], T3[13]
CV?	T4[55]	T5[12]	T4[55], T5[12]

Furthermore, when syllables are combined into words or phrases, lexical tones undergo sandhi changes in Shanghai Wu [3][5][25-28]. The general consensus is that given a sandhi domain, the tonal contours of non-initial syllables never surface, and the tone of the first syllable determines the f<sub>0</sub> contour of the whole domain. Specifically, in T1-T4 groups, the first syllable within the sandhi domain remains its base tone and assigns them to the following syllable, which is known as tone rightward spreading. In T5 group, the tonal contour of the first syllable is shifted to the final syllable and the previous syllable is assigned with [L] tone, which is known as tone shifting rule [27, 28]. The tone sandhi rules of the

disyllabic groups are illustrated in Table 2. As we can see, in disyllabic groups, the f0 contour of the first syllable (S1) becomes high-level (T1, T2, T4) or low-level (T3, T5), reflecting its register feature; the base tone contrast of the second syllable (S2) is neutralized and its tonal target is assigned by the base tone of S1. Take T3+X group for example, the f0 contour of S2 is high level, no matter its base tone is high register or low register. In other words, the co-occurrence restrictions between tone and onset consonant no longer holds. Here arise our research questions:

(1) In a disyllabic group, if the f0 contour of S2, which is determined by the phonological tone sandhi rules, is incompatible with its base tone (or its corresponding onset consonant), would the base tone of S2 affect the f0 implementation of sandhi tone? In other words, would the consonant-f0 interaction take effect during the tone sandhi process?

(2) If so, would the influence be attributed to purely automatic results of consonant production (i.e. consonant-f0 interaction) or be due to the phonetic implementation of the phonological feature contrast (i.e. register contrast)? If it is purely caused by the phonetic production of consonants, the perturbation would be found only in the syllable with a plosive onset, otherwise, it would appear in both consonant types.

(3) According to the previous studies on the f0 perturbation effect in Standard Chinese [23] and Cantonese [6], f0 perturbation is conditioned by the f0 level in that lower-range f0 accommodating greater magnitude and/or greater extent of consonant-f0 perturbation. Would the pitch context (the tonal targets of S1 or S2) affect the f0 perturbation effect in Shanghai Wu?

**Table 2:** Tone sandhi rules in Shanghai Wu

Register	Tonal group	Sandhi tone	Sandhi rules
High	T1[53]+X	55+31	Tone spreading
	T2[34]+X	33+44	
	T4[55]+X		
Low	T3[13]+X	22+44	Tone shifting
	T5[12]+X	11+13	

## 2. METHODOLOGY

### 2.1. Stimuli

The stimuli consist of (1) 12 mono-syllables with three lexical tones (T1[HM], T3[LM] and T5[LMq]) and two consonant types (plosives and nasals), as shown in Table 3; and (2) 24 disyllabic compound nouns [modifier+noun], including 6 tonal combinations (T1+T1, T1+T3, T3+T1, T3+T3, T5+T1 and T5+T3) and 4 types of consonant

combinations ([plosive+plosive], [plosive+nasal], [nasal+plosive] and [nasal+nasal]), as shown in Table 3. These stimulus words were embedded in the carrier sentence: /ŋu<sup>3</sup> kaŋ<sup>2</sup> X gəŋ<sup>5</sup> gəŋ<sup>2</sup> zɪ<sup>3</sup>/ (I said X this word)

**Table 3:** Stimuli

		T1 group		T3 group		T5 group	
Mono-	Nasal	猫 /mɔ <sup>1</sup> /	妈 /ma <sup>1</sup> /	毛 /mɔ <sup>3</sup> /	马 /mɔ <sup>3</sup> /	木 /mɔ <sup>2</sup> /	袜 /ma <sup>2</sup> /
		cat	mum	hair, fur	horse	wood	socks
	Plosive	包 /pɔ <sup>1</sup> /	爸 /pa <sup>1</sup> /	鲍 /bo <sup>3</sup> /	桃 /dɔ <sup>3</sup> /	薄 /bo <sup>2</sup> /	笛 /di <sup>2</sup> /
		a surname	dad	a surname	peach	thin	flute
Di-	X+High tone	猫妈	猫爸	马妈	马爸	木妈	木爸
		/mɔ <sup>1</sup> ma <sup>1</sup> /	/mɔ <sup>1</sup> pa <sup>1</sup> /	/mɔ <sup>3</sup> ma <sup>1</sup> /	/mɔ <sup>3</sup> pa <sup>1</sup> /	/mɔ <sup>2</sup> ma <sup>1</sup> /	/mɔ <sup>2</sup> pa <sup>1</sup> /
		包妈	包爸	鲍妈	鲍爸	薄妈	薄爸
	X+low tone	/pɔ <sup>1</sup> ma <sup>1</sup> /	/pɔ <sup>1</sup> pa <sup>1</sup> /	/bo <sup>3</sup> ma <sup>1</sup> /	/bo <sup>3</sup> pa <sup>1</sup> /	/bo <sup>2</sup> ma <sup>1</sup> /	/bo <sup>2</sup> pa <sup>1</sup> /
		猫妹	猫弟	马毛	马桃	木毛	木桃
		/mɔ <sup>1</sup> me <sup>3</sup> /	/mɔ <sup>1</sup> di <sup>3</sup> /	/mɔ <sup>3</sup> mo <sup>3</sup> /	/mɔ <sup>3</sup> do <sup>3</sup> /	/mɔ <sup>2</sup> mo <sup>3</sup> /	/mɔ <sup>2</sup> do <sup>3</sup> /
	胞妹	胞弟	鲍毛	鲍桃	薄毛	薄桃	
	/pɔ <sup>1</sup> mo <sup>3</sup> /	/pɔ <sup>1</sup> do <sup>3</sup> /	/bo <sup>3</sup> mo <sup>3</sup> /	/bo <sup>3</sup> do <sup>3</sup> /	/bo <sup>2</sup> mo <sup>3</sup> /	/bo <sup>2</sup> do <sup>3</sup> /	

### 2.2. Subjects and Recording procedure

4 male speakers between 25 to 35 years old, born and raised in Shanghai urban areas, participated in the study. They were paid for their participation and none of them reported any hearing, vision, or reading deficiencies.

The recording was conducted in the soundproof booth at Tongji University. Participants were seated in front of the computer and were asked to read aloud the target sentence naturally. The sentences were presented in PowerPoint slides in random orders. Every participant read the material three times. So, we got 36 words\* 4 speakers \*3 times=432 tokens.

### 2.3. Labelling and Measurements

The acoustic analysis was performed in Praat [2]. The onset and offset of vowels were manually labelled by referring to the change of F2 in the spectrogram. Then the Praat script ‘‘ProsodyPro’’ [22] was run and automatically provided us with accurate f0 tracks using a method that combines automatic vocal pulse marking by Praat, a trimming algorithm that removes spikes and sharp edges and a triangular smoothing function (cf. Appendix 1 in [23]).

Subsequently, the f0 measurements in Hz were converted to semitone relative to 50 Hz using the formula in (1) to better reflect pitch perception [21]. Formula (1) relates frequency in semitones, F, to frequency in Hz, f:

$$F=12*\log_2(f/50) \quad (1)$$

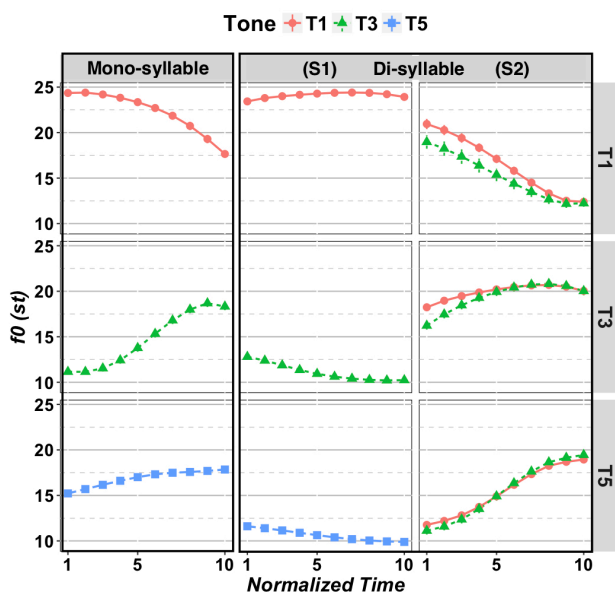
Linear Mixed-Effects models (using the lme4 package) were used to investigate how f0, duration and intensity were affected by syllable position, base tone and consonant type. All statistical analyses were carried out in R version 3.3.2 [20] using the lme4 package version 1.1-12 [1].

### 3. RESULTS

#### 3.1. General description

The time-normalized f0 contours (with 10 equidistant points) for mono-syllable (the left column) and disyllabic words (the two columns on the right) are given in Figure 1. The f0 contours within each row are organized by the same base tone of the initial syllable. Taking the first row for example, the f0 contour of the first column is the time-normalized f0 of mono-syllable with T1, averaged across 4 stimuli, 4 speakers and 3 repetitions. The f0 contours of the right column are the time-normalized f0 of the disyllabic groups, including 2 tonal combinations: X+T1 (red), X+T3 (green).

**Figure 1:** The time-normalized f0 contours of mono-syllable and disyllabic groups. S1 and S2 stand for the 1st and 2nd syllable of the disyllabic groups respectively. “Tone” stands for the base tone of each syllable, red for T1, green for T3 and blue for T5.



From Figure 1, we can see that in mono-syllable, T1 is a high-falling tone, starting from the top of the tonal register and falling until the middle of the register. T3 and T5 are both low-rising tones, but the f0 rising slope and f0 excursion size of T3 are larger than those of T5. In the disyllabic groups, tone sandhi is undertaken as expected. Within each group, the f0 contours of S1 become high-level or low-level, depending on their base tone register; while the f0 contours of S2 converge together, regardless of their different base tones. Specifically, in T1 group, the f0 contour of S1 becomes high-level and that of S2 becomes high-falling. In T3 group, the f0 contour of S1 becomes low-level and the f0 contour of S2 is high level with a rise at the beginning part. In T5 group,

the f0 contours of S1 and S2 become low-level and low rising, respectively.

Within the disyllabic group, although S2 loses its base tone contrast and their f0 contours converge together, there are differences in the onset f0 of S2. Specifically, the onset f0 of S2 with high base tone (i.e. T1) is higher than those with low base tone (i.e. T3), which indicates that although the tonal targets of S2 are assigned by the base tone of S1, the register contrast is partially retained and is reflected through the onset f0 of S2. Furthermore, the difference magnitude is larger in T1 and T3 groups, but is quite small in T5 group.

#### 3.2. Statistical analysis

In order to investigate the link between phonetics and phonology during the tone sandhi process, the statistical analyses were carried out in two steps.

First, the linear mixed effects models were constructed on the onset f0 and offset f0 of each syllable in disyllabic groups respectively, with the base tone of S1, S2 and their interaction as fixed factors, and with repetition and subject as random factors. The results are summarized in Table 4, which shows that:

With regard to S1, the base tone of S1 has significant effects on the onset f0 and offset f0, while the base tone of S2 has no significant effects. Specifically, the onset and offset f0 of T3 and T5 are significantly lower than those of T1, and the offset f0 of T3 and T5 is even lower than their onset f0, which confirms our observation above that the f0 contours of S1 in T3 and T5 groups are low slightly falling while that of T1 is high level.

With regard to S2, the base tone of S1 has significant effects on its onset f0 and offset f0, which indicates that in a disyllabic group, the f0 contour of S2 is determined by the base tone of S1. Comparing to T1 group, the onset f0 of S2 in T3 and T5 groups is significantly lower, especially in T5 group; while the offset f0 of S2 in T3 and T5 groups is significantly higher. This result confirms our observation above that the f0 contour of S2 is high-falling in T1 group, is high-level with a rising part at the beginning in T3 group, and is low rising in T5 group. Furthermore, the base tone of S2 had significant effects on the onset f0 of S2, as the onset f0 of S2 with a low base tone (i.e. T3) is significantly lower than that with a high base tone (i.e. T1), which echoes the results in Chen [4]. Furthermore, although the interaction of the two base tones (i.e. tones of S1 and S2) is not significant, we still can see that the lowering effect induced by the low register of S2 is less significant in T5 group, if comparing to that in T1 and T3 groups, which confirms our observation above.

Such results indicate that although the base tone contrast of S2 is neutralized by tone sandhi process, the pitch register contrast of S2 is partly maintained and reflected on the onset f0. However, the f0 perturbation is caused by the pure phonetic production of consonants or is due to the phonetic realization of phonology feature, which will be investigated in the following section.

**Table 4:** The results of onset and offset f0 in both syllables in disyllabic groups

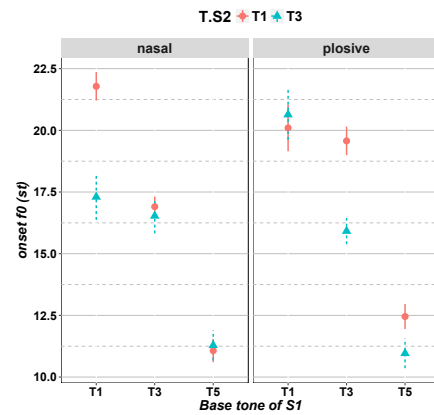
S1	onset f0				offset f0			
	Estimate	Std. E	t	p	Estimate	Std. E	t	p
(Intercept)	22.987	0.661	34.758	0.000	24.079	0.676	35.619	0.000
S1(T3)	-11.016	0.431	-25.576	0.000	-13.877	0.489	-28.394	0.000
S1(T5)	-11.709	0.431	-27.186	0.000	-14.344	0.489	-29.349	0.000
S2(T3)	-0.300	0.431	-0.697	0.486	-0.211	0.489	-0.433	0.665
S1(T3):S2(T3)	0.776	0.609	1.274	0.203	0.388	0.691	0.562	0.574
S1(T5):S2(T3)	-0.218	0.609	-0.357	0.721	0.632	0.691	0.915	0.360
S2	Estimate	Std. E	t	p	Estimate	Std. E	t	p
(Intercept)	20.545	0.809	25.385	0.000	12.412	0.715	17.370	0.000
S1(T3)	-2.704	0.538	-5.029	0.000	7.650	0.319	23.997	0.000
S1(T5)	-9.180	0.538	-17.072	0.000	6.549	0.319	20.545	0.000
S2(T3)	-1.973	0.538	-3.669	0.000	-0.161	0.319	-0.504	0.614
S1(T3):S2(T3)	-0.041	0.760	-0.054	0.957	0.140	0.451	0.311	0.756
S1(T5):S2(T3)	1.337	0.760	1.758	0.079	0.667	0.451	1.480	0.139

Second, in order to determine the cause of the onset f0 perturbation and the effects of the pitch context on it, the onset f0 data of S2 was divided according to consonant type (nasal vs plosive). Then a series of linear mixed effects models were run with base tone of S2, S1 and their interaction as fixed factors, and with repetition and subject as random factors. The results are illustrated in Figure 2, which shows that:

In both groups with nasal or plosive consonants, the register contrast of S2 is reflected to some extent, but in different pitch contexts. As for the group with nasal consonants, the register contrast of S2 is fully reflected in T1 group, as the onset f0 of S2 with T3 and T5 is significantly lower than that with T1; while there is no significant difference in T3 and T5 groups. In the contrary, the register contrast in plosive group is reflected in T3 and T5 groups, while there is no significant difference in T1 group. Such results indicate two things. First, the onset f0 perturbation is not caused purely by the production of voiced consonants, as there is the same pattern in the nasal group where there is no consonant contrast (voiceless vs voiced); rather it is caused by the phonetic implementation of the phonological feature, i.e. register contrast. Speakers do voluntarily control their articulatory settings, at least to some extent, so as to enhance the register contrast of S2. Second, the f0 perturbation pattern in plosive group is in line with the previous studies in Standard Chinese and Cantonese, in that consonant-f0 perturbation has a

great magnitude and longer extension in lower-range f0 environment, i.e. T3/T5 group in our study. However, the f0 perturbation pattern in nasal group shows an opposite trend, in that it is more significant in higher-range f0 environment, i.e. T1 group. Obviously, the f0 perturbation in nasal group is not caused by consonant-f0 interaction, and that's why they have the opposite pattern from the plosive group. However, further reasons still need more investigation.

**Figure 2:** The effects of the register of S2 on the onset f0 in different tonal groups.



#### 4. DISCUSSION AND CONCLUSION

We investigated the effects of base tone and consonant type of the second syllable (S2) on the f0 implementation of tone sandhi in Shanghai Wu, in order to further our understanding of the interaction of phonetic mechanism and phonological rule during the tone sandhi. Results showed that although the base tone contrast of S2 is neutralized by sandhi rule in Shanghai, the register contrast of its base tone is retained and is reflected by the onset f0, as the onset f0 of S2 with low tones is significantly lower than that with high tones. This onset f0 perturbation is not caused purely by the phonetic production of consonants, as voiced onsets tend to lower the f0 of the following vowel; rather it is due to the phonetic realization of the phonological feature, i.e. register contrast, because we observe the same pattern in the syllable with the same nasal onset /m/. Speakers intentionally manipulate phonetic cues (i.e., f0) to enhance the perception of the base tone/register contrast of S2, in order to help listeners to distinguish the morphemes with the same segments but different base tones, as the f0 contour no longer a reliable cue in S2. Therefore, the f0 perturbation in non-initial positions is phonologically controlled, i.e., intended by speakers who actively adjust their articulatory settings to enhance/signal certain phonological contrasts in the language.

## 5. REFERENCES

- [1]. Bates, D., Maechler, M., Bolker, B., & Walker, S., 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-12.
- [2]. Boersma, P., & Weenink, D., 2010. Praat: Doing phonetics by computer (Version 5.1.30) [Computer program]. Retrieved from <http://www.praat.org>
- [3]. Chen, Y. Y. "Revisiting the phonetics and phonology of tone sandhi in Shanghai Wu", *Speech Prosody 2008*, Campinas, Brazil, 253–256.
- [4]. Chen, Y. Y. 2011. How does phonology guide phonetics in segment–f0 interaction? *Journal of Phonetics*, 39, 612–625.
- [5]. Duanmu, S. "Rime length, stress, and association domains", *Journal of East Asian Linguistics*, vol. 2, pp: 1–44, 1993
- [6]. Francis, A. L., Ciocca, V., Wong, V. K. M., & Chan, J. K. L. 2006. Is fundamental frequency a cue to aspiration in initial stop?. *Journal of the Acoustical Society of America*, 120, 2884–2895.
- [7]. House, A. S., & Fairbanks, G. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, 25, 105–113.
- [8]. Hombert, J.-M. 1978. Consonant types, vowel quality, and tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 77–111). New York: Academic Press.
- [9]. Hyman, L. (Ed.) 1973. *Consonant types and tone: Southern California occasional papers in linguistics*. Los Angeles: The Linguistic Program, University of Southern California.
- [10]. Halle, M., & Stevens N. 1971. A note on laryngeal features (pp. 45–61). Berlin: Mouton de Gruyter (Reprinted in M. Halle, *From memory to speech and back: Papers on phonetics and phonology*, 2002).
- [11]. Honda, K. 2004. Physiological factors causing tonal characteristics of speech: From global to local prosody. In *Proceedings of the Speech Prosody 2004*, Nara, Japan.
- [12]. Hoole, P. 2006. *Experimental studies of laryngeal articulation*. Germany: Habilitationsschrift, University of Munich.
- [13]. Hombert, J., Ohala, J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, 55, 37–58.
- [14]. Jun, S. A. 1996. Influence of microprosody on macroprosody: A case study of phrase initial strengthening. *UCLA Working Papers in Phonetics*, 92, 97–116.
- [15]. Kingston, J., & Diehl, R. 1994. Phonetic knowledge. *Language*, 70, 419–454.
- [16]. Kohler, K. J. 1982. F0 in the production of lenis and fortis plosives. *Phonetica*, 39, 199–218.
- [17]. Lehiste, I., & Peterson, G. E. 1961. Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America*, 33, 419–425.
- [18]. Lofqvist, A., Baer, T., McGarr, N., & Story, R. S. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America*, 85, 1314–1321.
- [19]. Ohala, J. 1978. Production of tone. In V. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 5–39). New York: Academic Press.
- [20]. R Core Team, 2016. R: A language and environment for statistical computing, Vienna, Austria: R Foundation for Statistical Computing (Version 3.3.2).
- [21]. Rietveld, T., & Chen, A.-J., 2006. How to obtain and process perceptual judgements of intonational meaning. In S. Sudhoff, D. Lenortová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, & J. Schießer (Eds.), *Methods in empirical prosody research* (pp. 283–319). Berlin: Walter de Gruyter.
- [22]. Xu, Y. 1999. Effects of tone and focus on the formation and alignment of F0 contour, *Journal of Phonetics*, 27, 55–105.
- [23]. Xu, C. X., & Xu, Y. 2003. Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Alphabet*, 33, 165–181.
- [24]. Xu, Y., 2013. *ProsodyPro — A tool for large-scale systematic prosody analysis*. Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013). Aix-en-Provence, France.
- [25]. Xu, B. H., Tang, Z. Z., & Qian, N.R. 1981. Tone sandhi in New Shanghainese, *Dialects*, 3, 145-155.
- [26]. Yip, M. 2002. *Tone*. Cambridge: Cambridge University Press.
- [27]. Zhu, X.N. 2005. *An Experimental Study in Shanghai Tones*. Shanghai: Shanghai Educational Publishing House.
- [28]. Zee, E., & Maddieson, I., 1980. Tones and tone sandhi in Shanghai: Phonetic evidence and phonological analysis. *Glossa*, 14, 45–88.