A Study on Mandarin Chinese "Bu" Tone Sandhi Followed by English Words

Kaige Gao[1] and Xiyu Wu[2]

Laboratory of Linguistics, Department of Chinese Language and Literature, Peking University

[1]michellegkg@pku.edu.cn, [2]xiyuwu@pku.edu.cn

Abstract

This paper focuses on tone sandhi of the Chinese word "Bu" (不) followed by English words. In Mandarin Chinese, the tone of the word "Bu" is tone 4 (the high-falling tone), and its tone sandhi rule is that it becomes tone 2 (the mid-rising tone) when followed by another tone 4 syllable. However, few researchers have paid attention to the tone sandhi rules of "Bu" when followed by English words. In this study, the tone sandhi rule of "Bu" when followed by English words is explored by a perceptual experiment and a pitch contour analysis. Results show that the Mandarin Chinese "Bu" tone sandhi also applies when "Bu" is followed by an English word. When followed by a high-falling monosyllabic English word, "Bu" will become tone 2. Furthermore, the pitch contour of English words embedded in Mandarin sentences is predictable according to their syllabic structure and lexical stress. This paper further proves that native Chinese speakers perceive English words inserted into Chinese speech as tonal-like language. The tonal patterns of English words summarized in this study can provide theoretical support for improving the naturalness of mixed-language speech synthesis.

Index Terms: tone sandhi, mixed-language, "Bu"

1. Introduction

The demand for multilingual capability increases as global communication grows and inserting English words into Mandarin Chinese has been very common in recent years. In practical use, making the mixed-language speech natural and fluent has been a thorny issue in TTS (text-to-speech). Some methods have been proposed to generate more natural speech sounds. For example, Kuo et al. treated each English word as a Chinese word to generate prosodic parameters for each alphabet letter of the word[1]. Lai proposed a pitch-modeling approach for Mandarin Chinese Speech mixed with English letters[2]. However, these studies focus on English abbreviation word spelling such as "KTV" and "BBS" that don't share the pitch model with real English words. Moreover, the effect of embedded English words on the tone sandhi of Mandarin Chinese has not received enough attention.

One of the main issues in synthesizing mixed-language speech is how words of non-tonal languages, when entering a tonal language, are produced and perceived by native speakers of the tonal language. The way words of the non-tonal language are produced and perceived may directly affect the tone sandhi in the tonal language. In this paper, we take the mixing of Chinese and English as an example and focus on the tone sandhi of the word "Bu" followed by an English word, which is common and representative.

There are four tones on full syllables in Mandarin Chinese, namely tone 1, tone 2, tone 3, and tone 4, whose canonical forms are high-level tone, mid-rising tone, dip-rising tone, and high-falling tone, shown in Table 1[3].

Table 1: *Four tones in Mandarin Chinese marked with Chao tone numerals*

	Tone 1	**Tone 2**	**Tone 3**	**Tone 4**
Phonetic Form	ma55	ma35	ma214	ma51
Gloss	'mother'	'hemp'	'horse'	'scold'

Tone sandhi is a phonological process that often occurs in Mandarin Chinese. The word "Bu", which means 'not' in Chinese and is generally used in negative constructions, is typical. The tone sandhi rule of "Bu" is that it remains tone 4 before tone 1, tone 2, and tone 3, and it becomes tone 2 before another tone 4[3]. Here is an example of its tone sandhi.

E.g. 不对(bu51 + dui51 → bu35 dui51)

In brief, the tone sandhi of "Bu" relies on the tone of the following syllable. The rule is well observed when converting pure Mandarin Chinese text to speech, but there is a problem with the naturalness when "Bu" is followed by an English word, a word of a non-tonal language. For example, when converting the text "不care" to speech, the word "Bu" usually remains tone 4 and does not change to tone 2 in the current commercial TTS system, which sounds stiff and does not match the native speaker's intuition. Obviously, despite the fact that tone sandhi rule of "Bu" has been well studied in pure Mandarin, it is not clear how the insertion of English words affects it. Therefore, the goal of this paper is to find out the rule of tone sandhi when "Bu" is followed by an English word in Mandarin Chinese, which is worth investigating because it helps both to synthesize more natural mixed-language speech and to reveal the nature of tone sandhi.

It should be noted that the phenomenon we focus on is different from code-switching and transliteration. Cheng first described the differences[4]. Firstly, in these cases the English words are used in the Chinese sentences not because of a lack of certain words in Chinese. For example, there are Chinese words for "care" (在乎), "cancel" (取消), and "man" (男人). But some Chinese speakers tend to insert English words into Chinese due to their speech habits. Besides, compared with code-switching, English is simply inserted into the sentences in the form of words, with no grammatical dimension involved. Furthermore, they do not borrow English words into their language system; rather, they use English words directly. In other words, the English words inserted in sentences retain their syllabic structure instead of turning into transliteration words like "沙发(sha55 fa55)"(transliteration of sofa), "巴士(ba55 shi51)"(transliteration of bus), etc.

Several perceptual studies have provided the theoretical basis for this study. Researchers have revealed that native Chinese speakers are very sensitive to tonal information and maintain this sensitivity in a multilingual context. Shook and Marian provide evidence for bilinguals' sensitivity to tonal information, even when listening to a non-tonal language like English[5].

979-8-3503-9796-3/22/$31.00

Ortega-Llebaria and Wu prove that Chinese-English speakers reinterpret English as a tonal-like language and continue processing pitch at the lexical level in both Chinese and English[6]. In this study, if the embedded English words affects the tone sandhi of "Bu", it helps to further demonstrate that the inserted English words are perceived as tonal by native Chinese speakers.

Based on previous work and the tone sandhi rule of the word "Bu" in Mandarin Chinese, we propose three questions as follows.

- Firstly, will "Bu" change its tone depending on the English word that follows?
- Secondly, if it will, does "Bu" become tone 2 before an English word whose first syllable is in a descending pitch?
- Lastly, is there a specific pattern in the pitch contour of the first syllable of English words when embedded into Mandarin Chinese?

2. Experiment

2.1. Data acquisition

A Mandarin-Chinese speech database is used in this study, which consists of 28 different declarative sentences[1] , each containing "*'Bu' + An English word*" , e.g., 你提的这个问题一点都不*low*，蛮有价值的。*(Your question is not low at all, it is quite valuable.)* The English words are inserted into sentences to ensure the utterances are as natural as possible. The 28 English words have maximum coverage of consonant phonemes that can be used as the beginning of a word in English except /z/ and /ð/. Because English words that can be inserted after "Bu" are usually verbs and adjectives and there are relatively few verbs and adjectives that begin with these two consonants.

Four native speakers of Mandarin Chinese, two males, and two females, all of whom are students at Peking University, generated all utterances. They had no long-term experience living in an English-speaking environment and all passed CET-6 (College English Test-6). All speech signals were digitally recorded by the Audition software and sampled in mono at 20-kHz. All sentences were read naturally at an average speed of 3.5 syllables per second. Each sentence was read twice.

2.2. Parameter analysis

The parameter analysis consists of two main parts.

Firstly, to confirm the occurrence of tone sandhi and determine before which words "Bu" are pronounced as tone 2, we conducted a perceptual experiment. A total of 12 native speakers of Mandarin Chinese at Peking University (6 female, 6 male) were recruited to conduct a perceptual experiment on the corpus of four speakers. Each subject was asked to listen to 28 sentences pronounced by one of the speakers mentioned above, and each sentence was played twice. That is, each speaker's utterances were heard by three subjects. After listening to the utterances, the subjects were asked to choose whether the tone of the word "Bu" in the sentence was tone 2, tone 4, or indeterminate. Then we calculated the consistency rate of 12 subjects on each sentence. For example, if all subjects perceived the tone of "Bu" before the word "low" as tone 2, then the consistency rate of this word is one hundred percent. Besides the experiment above, we also checked the pitch contour of "Bu", especially whose consistency rate did not reach 90% in the experimental results.

Secondly, the fundamental frequency of the first syllable of each English word in the corpus is extracted. After converting the fundamental frequency to semitones, the pitch of the first syllable of each English word is annotated using the Chao tone numerals, which divides the pitch range into 5 degrees, and uses five numbers from 1 to 5 to indicate the pitch and direction of the tone where 5 indicates the highest and 1 the lowest[7]. We adopt the Chao tone numerals because it is usually used to mark Chinese tones. Since we assume that English words are also toned when embedded in Chinese, the similarities and differences between English word tones and Chinese syllable tones can be seen more clearly using the same marking method.

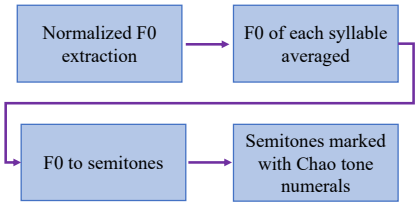

Figure 1: *Diagram of converting F0 to Chao tone numerals*

Our pipeline is shown in Figure 1. We first extract the fundamental frequency of the first syllable of each English word as follows,

$$f_0^i = g_{Kawahara}(x^i), \tag{1}$$

where x^i and f_0^i represent the input audio and its corresponding fundamental frequency of the speaker's i-th attempt. $g_{Kawahara}(\cdot)$ is the fundamental frequency extraction algorithm proposed by Kawahara[8]. After that, the length of time is normalized as follows,

$$\hat{f}_0^i = g_{normalize}(f_0^i; N), \tag{2}$$

where $g_{normalize}(\cdot;\cdot)$ uniformly samples N points from f_0 to unify sequence length. Here we set N as 20. The results are averaged for each speaker reading the same word. We note this procedure as

$$\hat{f}_0 = \frac{\sum_{i=1}^{M} \hat{f}_0^i}{M}, \tag{3}$$

where M represents the times the speaker repeats the text. In this study, M equals 2. Then the averaged value of each syllable is converted to a semitone value, which is shown in eq. 4.

$$st = 12log_2(\frac{\hat{f}_0}{f_{ref}}). \tag{4}$$

Here f_{ref} is the minimum of the fundamental frequency of each speaker's English pronunciation. The lowest semitone of the fundamental frequency is converted to zero. After averaging the range of semitones into fifths, the pitch contour of the first syllable of each English word is marked using Chao tone numerals by applying the eq. 5:

$$tn = \lceil \frac{st}{interval} \rceil. \tag{5}$$

$interval$ is set as $\frac{st_{max}}{5}$, where st_{max} is the maximum of st. After completing the above steps, the pitch of each syllable is

[1]The sentences and the speech samples are given in *https://docs.google.com/document/d/1llM7ctDpTjyRW41m8tyR8bdBjZpFsa06NopniPXoBxE/edit?usp=sharing*

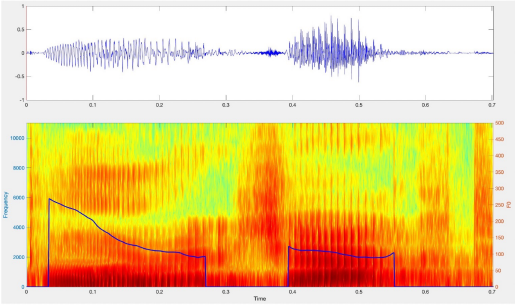

Figure 2: *Spectrogram of "Bu51 reject"*

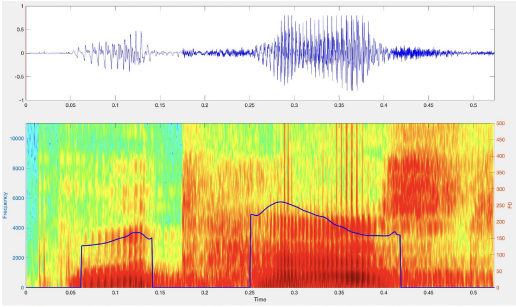

Figure 3: *Spectrogram of "Bu35 test"*

marked by 20 numerals with values from 1 to 5. Based on the distribution of these 20 numerals, we determine the pitch contour of each syllable.

3. Results

Based on the perceptual experiment and the pitch contour analysis, it can be concluded that four speakers almost change the tone of "Bu" in the same way. Table 2 shows that the tone of "Bu" in front of different English words is different. The tone of "Bu" remains tone 4 in front of argue, boring, call, cancel, delicious, fashion, generous, google, hip-hop, low, man, polite, reject, satisfied, solid, thirsty, visit, and becomes tone 2 in front of care, chill, cool, fair, nice, push, sure, get, test, work and young. Moreover, the consistency rate is over ninety percent in most cases, indicating that at least eleven of the twelve subjects gave consistent responses, and the tone sandhi of "Bu" followed by English words is not random but regular. In addition to the perception experiment, we also checked the tone of "Bu" based on the pitch contour, especially for those without a one hundred percent consistency rate. For instance, the spectrograms in Figure 2 and Figure 3 show the pitch contours of "Bu51 reject" and "Bu35 test". From these two figures, one can easily see that the pitch contour of "Bu" varies. The tone of "Bu" remains high-falling in front of "reject" while becomes the rising tone when followed by "test".

To have a better understanding of the relationship between the pitch patterns of the initial syllable of English words and their syllabic structure, these 28 words are divided into three categories according to the pitch contour of their first syllable as shown in Table 3, including level tone, high-falling tone, and low-falling tone. It should be noted that there are a little bit differences in the tone numerals within each category. The classification criteria are as follows. Expressed in Chao tone numerals, the type of level tone includes 22, 33, 44, and 55, where 33 and 44 are the most common. The type of high-falling tone includes 31, 41, 42, 51, 52, 53. That is, we classify any tone whose beginning and ending differ by more than 2 degrees as a high-falling tone. The type of low-falling only includes 32, which starts low and the downward trend is not significant.

Words in Table 3 marked with an asterisk including get, test, and work, whose pitch contours are not always strictly high-falling will be discussed in the next section.

Table 2: *Tone of "Bu" before words and the Consistency rate derived from the perceptual experiment*

The tone of "Bu"	**Words and the Consistency rate**
tone 2	care(75%), chill(100%), cool(100%), fair(92%), get(100%), nice(100%), push(100%), sure(92%), test(92%), work(100%), young(100%)
tone 4	argue(100%), boring(100%), call(92%), cancel(83%), delicious(100%), fashion(100%), generous(100%), google(100%), hiphop(92%), low(100%), man(83%), polite(100%), reject(75%), satisfied(92%), solid(67%), thirsty(92%), visit(100%)

4. Discussions

It is basically certain that the tone sandhi of "Bu" is influenced by the following English word so the answer to the first question is: yes, "Bu" changes its tone depending on the English word that follows.

When it comes to the second question, does "Bu" become tone 2 before an English word whose first syllable is in a descending pitch? Not all descending pitch contours will cause "Bu" to change its tone. As shown in Table 3, when embedded into Mandarin Chinese speech, there are three types of pitch contour of the first syllable of the English words: level, high-falling, and low-falling. It should be noted that when followed by a high-falling syllable, the tone of "Bu" is different from that followed by a low-falling syllable. The tone of "Bu" remains tone 4 when the first syllable of an English word is level or low-falling, while it becomes tone 2 when the first syllable is high-falling. Therefore, the rule of tone sandhi of "Bu" is that when followed by a high-falling monosyllabic word, "Bu" will become tone 2.

As for the last question, is there a specific pattern in the pitch contour of the first syllable of English words when embedded into Mandarin Chinese? Generally speaking, when reading an English word embedded in Mandarin, Chinese-English speakers produce the first syllable with a fixed pitch pattern, which is relevant to the word's syllabic structure and lexical stress. From the data collected in this study, it can be concluded that monosyllabic words exhibit either high-falling or level tones. Polysyllabic words tend to have a level or low-falling initial syllable when inserted into Chinese speech. Specifically, the pitch contour of the first syllable is level when the lexical stress falls on the initial syllable, and it is low-falling when the lexical stress is not in the initial syllable. The pattern is shown in Table 3.

However, some monosyllabic words don't fully conform to the rule above. Pitch contours of some monosyllabic words including get, test, and work are not always strictly high-falling.

Table 3: *The pitch contour pattern of the words' first syllable*

Syllabic structure	Lexical stress	Pitch contour of the initial syllable	Words	Tone of "Bu"
monosyllabic	front	high-falling	care, chill, cool, fair, nice, push, sure, *get, *test, *work	tone 2
		level	call, low, man	tone 4
polysyllabic	front	level	argue, boring, cancel, fashion, generous, google, hiphop, reject, satisfied, solid, thirsty, visit	tone 4
	non-front	low-falling	delicious, polite, reject	tone 4

Table 4: *Pitch contour marked with Chao tone numerals of get, test and work*

	get	test	work
speaker1	54	54	54
speaker2	53	52	41
speaker3	54	54	43
speaker4	52	43	53

As shown in Table 4, in some cases the beginning and ending of these words don't differ by more than 2 degrees. They all start high and have a downward trend, but sometimes not completely down by the end of the plosive. Although these three words don't exhibit standard high-falling pitch acoustically, they are treated like the high-falling pitch by native speakers, for the tone of "Bu" in front will become tone 2 and the consistency rate of these words is over ninety percent. From these three words, the fact that their pitch does not descend to the bottom of the speaker's pitch range may be due to the presence of a final voiceless plosive, which usually results in insufficient time for the pitch to decline. In addition, it shows that there are acoustic and perceptual differences in the lexical tone of English words for native Chinese speakers.

Generally, native Chinese speakers prefer to encode a falling pitch in monosyllabic English words[6]. However, several monosyllabic words like man, low, and call are level instead of high-falling when inserted into Mandarin Chinese. A possible answer is that these words have fused with some Chinese morphemes and the fusion is usually in a high frequency of usage in everyday conversation.[9]. For example, it could be the case that "call" is often used in the chunking "打call"(to support someone), which is pretty common in Mandarin Chinese conversation. In this chunking, "call" is usually pronounced as a level tone, so "call" tends to be a level tone whenever inserted into Mandarin Chinese. However, no clear pattern has been identified as to which of the monosyllabic words will exhibit level tone, and further research is needed. At present, such words as "call" can be marked out separately as irregular words, since the number of monosyllabic words exhibiting the level tone is much less than those exhibiting the high-falling tone.

5. Conclusions and Future work

This paper presents the discovery that the tone sandhi of the word "Bu" in Mandarin Chinese also occurs when an English word follows it. We found that when followed by a high-falling monosyllabic word, "Bu" will become tone 2. Moreover, the pitch contour of English words pronounced by native Chinese speakers is related to their syllabic structure and lexical stress.

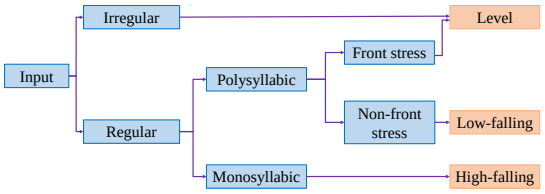

Figure 4: *Prediction process of the pitch contour of the initial syllable*

Figure 4 shows how to predict the pitch contour of an English word's first syllable. It can be concluded that monosyllabic words exhibit either high-falling or level tones while polysyllabic words tend to have a level or low-falling initial syllable when inserted into Chinese speech. In particular, for polysyllabic words, the first syllable is level when the lexical stress falls on the initial syllable, and low-falling when the lexical stress is not in the initial syllable. Importantly, this paper demonstrates that when Chinese and English are mixed, the tone sandhi of Chinese is also influenced by English, which is helpful to improve the naturalness of TTS, because the pattern we conclude can be a reference for synthesizing Chinese with inserted English words.

There are three directions for future work. Firstly, native Chinese speakers' perception of the tone of English words may differ from their acoustic pitch performance. For instance, monosyllabic words ending with a plosive are most likely to be perceived as high-falling, but their pitch contour is not standard high-falling. What are the specific factors that determine the perception of these words as the high-falling pitch? Is it the slope of the declination or the fundamental frequency at the beginning? Secondly, whether the monosyllabic English words that exhibit the level tone upon entering Mandarin Chinese are related to the frequency of use and common collocations needs to be further explored. To find out these monosyllabic words in level tone is beneficial to synthesizing more natural mixed-language speech. Lastly, it is worth studying whether it is possible to apply the tone sandhi rules to the mixture of Chinese and other non-tonal languages.

6. Acknowledgements

The Research would like to thank the supports of the National Social Science Foundation of China "Research on speech quantum theory based on physiological models" (project number: 18BYY189) and the major project of the Ministry of Education "Research on language ontology based on speech multimodality" (project number: 17JJD740001).

7. References

[1] W. C. Kuo, Y. R. Wang, H. M. Lu, and S. H. Chen, "An NN-based approach to prosody generation for English word spelling in English-Chinese bilingual TTS," in *International Symposium on Chinese Spoken Language Processing*, 2002.

[2] W. Lai, "Pitch modeling for Chinese Speech mixed with English word spelling," in *2008 9th International Conference on Signal Processing*. IEEE, 2008, pp. 592–595.

[3] S. Duanmu, *The phonology of standard Chinese*. OUP Oxford, 2007.

[4] C. C. Cheng, "English stresses and Chinese tones in Chinese sentences," *Phonetica*, vol. 18, no. 2, pp. 77–88, 1968.

[5] A. Shook and V. Marian, "The influence of native-language tones on lexical access in the second language," *The Journal of the Acoustical Society of America*, vol. 139, no. 6, pp. 3102–3109, 2016.

[6] M. Ortega-Llebaria and Z. Wu, "Chinese-english speakers' perception of pitch in their non-tonal language: Reinterpreting english as a tonal-like language," *Language and speech*, vol. 64, no. 2, pp. 467–487, 2021.

[7] Y. R. Chao, "A grammar of spoken Chinese," 1965.

[8] H. Kawahara, I. Masuda-Katsuse, and A. De Cheveigne, "Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based f0 extraction: Possible role of a repetitive structure in sounds," *Speech communication*, vol. 27, no. 3-4, pp. 187–207, 1999.

[9] I. Bybee, "Sequentiality as the basis," *The evolution of language out of pre-language*, vol. 53, pp. 109–134, 2002.