

IBM Research Report

Mandarin Chinese Tonal Issues from the Perspective of Speech Synthesis

Fangxin Chen

IBM Research Division
China Research Laboratory
HaoHai Building, No. 7, 5th Street
ShangDi, Beijing 100085
China

Jianfen Cao

Linguistics Institute
Chinese Academy of Social Science



Research Division

Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Mandarin Chinese Tonal Issues From The Perspective of Speech Synthesis

Fangxin Chen⁺, Jianfen Cao⁺⁺

⁺IBM China Research Laboratory

e-mail: chenfx@cn.ibm.com

⁺⁺Linguistics Institute, Chinese Academy of Social Science

e-mail: jianfencao@163.com.cn

Abstract

This paper presents two tonal issues in spoken Mandarin Chinese from the perspective of speech synthesis. One is a unique Chinese phonetic category *Qingsheng* (轻声). Based on speech synthesis and natural speech analysis, two acoustic criteria were suggested for distinguishing *Qingsheng* from the unstressed syllables which occur frequently in natural speech. The other is a tone *sandhi* phenomenon which differs from the known tonal behaviors and affected the application of normal tone *sandhi* rules in our text-to-Speech system. Tentative discussion was presented to solicit further theoretical enquiry.

1. Introduction

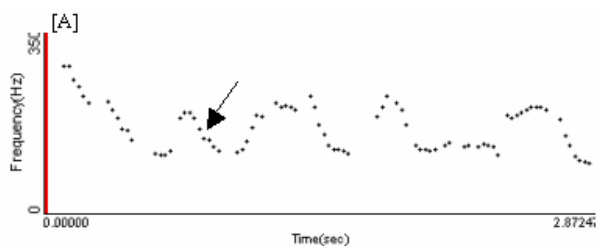
Linguistic theory has contributed significantly to speech technology. In turn, the latter has been stimulating new research interest in many linguistic areas. Recent years' surge in prosody investigation is an example, which was out of the interest in natural spoken language processing and modeling. The text-to-speech (TTS) work, on the other hand, has put traditional linguistic theories and assertions into scrutiny. In our Mandarin Chinese (MC) synthesis research, we have encountered various subtle phonetic issues not clearly defined, explained or accounted for by the traditional linguistic theories. Here we present two such tonal phenomena arising from MC synthesis to solicit more linguistic investigation and discussion.

2. Qingsheng

One unique linguistic phenomenon in spoken Mandarin Chinese(MC) is *Qingshen*, traditionally translated in English as *neutral tone syllable*, *atonic syllable*, etc.. *Qingsheng* is at syllable level and it is compulsory for a list of *Qingsheng* words in speech. A *Qingsheng* word could have different semantic meaning as compared with the word consisting of exactly the same Chinese Characters but read in normal lexical tones. Therefore, *Qingsheng* is a phonological issue.

In Chinese linguistics literature, there exists significant controversy in the *Qingsheng* concept. The major issue is how to view the tonal effect in *Qingsheng* perception. Some linguists considered it as a pure tonal phenomenon, while others argued that it was mainly a syllable stress problem. They even disagreed with the name of *Qingsheng*. Zhao[1] argued that *Qingsheng* was due to the reduction of duration and intensity level. The tonal effect was less significant. Therefore, *Qingyin* (轻音, weak voice) rather than *Qingsheng* should be used to avoid the Chinese character 声 (sheng)'s literal connotation with pitch. Based on her acoustic analyses and relevant synthesis tests, Cao[2] suggested that *Qingsheng* syllables had its own regular pitch pattern, which played a dominant role in *Qingsheng* perception

In our MC TTS experiment, a *Qingsheng* syllable was synthesized by applying the neutral tone pitch contours which depended on the lexical tone patterns of their preceding syllables [3], as well as reducing its duration (mainly the rhyme part) and energy level. The problem we encountered was that there were different versions of *Qingsheng/Qinyin* vocabulary existing in linguistic books. Some *Qingsheng/Qinyin* words did not sound natural when they were synthesized according to our *Qingsheng* model. One case was the word 讨论 (*Taolun*, discuss), which was given as a *Qingyin* word example in [1]. In synthesis of that word, because the preceding syllable *Tao* is a falling-rising tone (Tone3), the neutral tone in *lun* was implemented as a short mid-level pitch according to the neutral tone rule. But that pitch contour introduced unnaturalness to the word. For comparison, we examined the pitch contours of the same word in real speech by both male and female speakers.



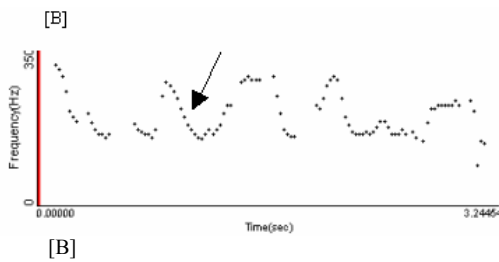


Figure.1. Pitch contours of the sentence 会上讨论了研究会十二年来的工作 spoken by [A] a male and [B] a female speaker.

The falling pitch part pointed by the arrow in both male and female speech illustrated by Figure 1 is the pitch contours for the syllable *lun*. As shown, the pitch contour of *lun* did not conform with the neutral tone pattern, but rather in its default high falling tone (*Tone4*). Based on observation on the real speech data, we re-synthesized the word 讨论 again by unstressing the syllable *lun*, which included reducing the syllable duration, energy level, pitch register and pitch range, etc., but the *Tone4* pattern was still kept. In other words, *lun* was reduced in all other acoustic aspects as a *Qingsheng* syllable, except that it retained its default lexical tone, *Tone4* in *lun*'s case. As a result, the re-synthesized word 讨论 became natural in perception. This indicates that *lun* in word 讨论 was only an unstressed syllable in context, rather than a *Qingsheng* syllable.

The synthesis result not only revealed the tonal effect in *Qingsheng* perception, but also suggested that the tonal pattern is the essential acoustic cue to separate *Qingsheng* from contextual unstressed syllable. The tonal pattern of the context-conditioned unstressed syllable is a simple lexical tone reduction phenomenon; while the tone of a *Qingsheng* syllable is not simply reduced, but also neutralized and resulted in pitch pattern.

Therefore, we suggest that, there should be two acoustic criteria for classifying *Qingsheng* vocabulary for MC. The first condition is that, the pitch contour in the syllable should conform with the neutral tone rules. If the preceding syllable is a high level tone (*Tone1*), rising tone (*Tone2*) or high falling tone (*Tone4*), the pitch of the *Qingsheng* syllable follows the pitch ending point of the preceding syllable and falls to some degree in the rest part of the syllable. If the preceding syllable is in *Tone3*, the pitch starting point in *Qingsheng* is raised above the end point pitch of the preceding syllable to certain degree and keeps relatively level, or falls slightly in the rest part of the syllable. The second condition is that, the syllable should be in unstressed status and its duration, energy and segment quality are all reduced.

The advantages of having the acoustic criteria for *Qingsheng* classification are that:

- There are concrete acoustic criteria for *Qingsheng* judgment. With that, the MC *Qingsheng* vocabulary could be constrained and standardized;
- The *Qingsheng* syllable is phonetically and acoustically separated from the unstressed syllable;
- Facilitate TTS system to apply different acoustic models to *Qingsheng* and unstressed syllables to get more natural synthesized speech.

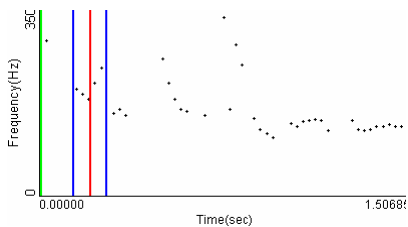
Currently there exist different English terms for *Qingsheng* based on different interpretations on it, such as *neutral tone*, *weak voice*, etc.. Those terms only presented one or the other aspect of *Qingsheng*. It seems better to use *Qingsheng* directly as a standard linguistic expression to avoid possible misinterpretation and misunderstanding of this MC phonetic phenomenon. Anyway, there already exist a number of specific terminologies in the linguistic vocabulary, such as *sandhi*, *Mora*, *Pinyin*, etc., it should be no problem by adding one more term *Qingsheng* to enrich the linguistic vocabulary.

3. New Tone Sandhi Phenomenon?

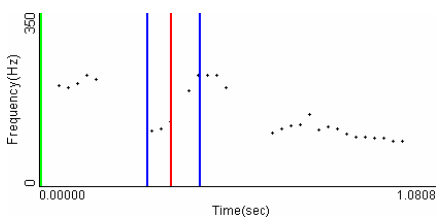
As known, MC has four lexical tones in citation form. Each tone has a context-dependent pitch pattern. In our TTS system, a number of lexical tone rules were implemented to accommodate the tone pattern variations arising from tone co-articulation, tone sandhi, syllable stress, word accent, sentence type etc..

In synthesis of the frequently used interrogative word 什么 (*Shenme*, meaning *what*), however, we encountered a puzzling problem when the normal lexical tone rules were applied to that word. *Shenme* is a *Qingsheng* word. The first syllable *shen* is in *Tone2*, and the second syllable *me* in neutral tone. Accordingly, the TTS system applied *Tone2* pattern to the syllable *shen*. In the syllable *me*, a downward tailing pitch contour was applied based on the neutral tone rule. However, the word synthesized that way sounded unnatural, which drove us to check the pitch patterns of the word in natural speech. Considering the possible tone co-articulation and tone *sandhi* effect, we selected sentences in our speech database in which the word 什么 was preceded by different lexical tone categories. The pitch contours between the first two bars in each graph in Figure2 belong to the syllable *shen*. The pitch contours between the second and the third bar to the syllable *me*.

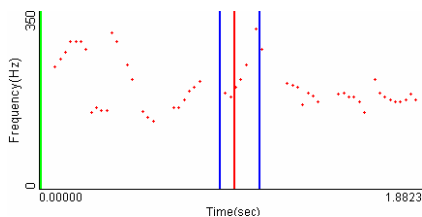
[a]. The word is at the sentence beginning position:
什么事这么着急?



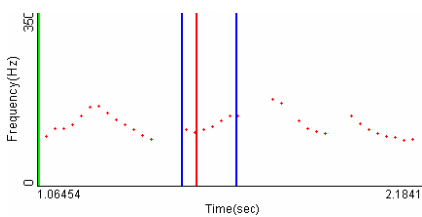
[b]. The word follows a syllable with Tone1:
都什么年代了



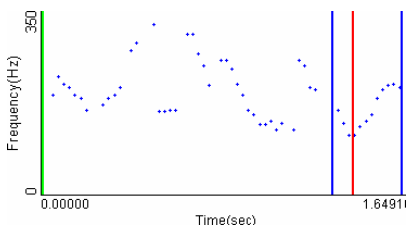
[c]. The word follows a syllable with Tone2:
明天早晨什么时候出发



[d]. The word follows a syllable with Tone3:
没有什么变化



[e]. The word follows a syllable with Tone4:
你还待在这里做什么?



[f]. The word follows a syllable with neutral tone:
别的什么都不管

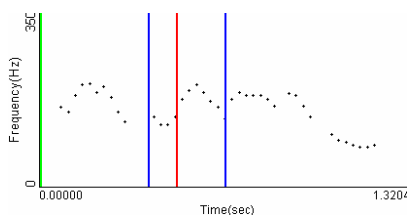


Figure 2. the pitch patterns of the word 什么 (*Shenme*) in different lexical tone contexts.

As shown in the above figure, the pitch contours of the syllable *shen* in most tone contexts did not fit into Tone2 pitch contours. The pitch was basically a low tone, which was significantly different from the normal Tone2 pitch pattern. Meantime, the pitch contours of the syllable *me* also did not fit into the neutral tone pattern. According to the neutral tone rule, a *neutral* tone when preceded by *Tone2*, supposedly should have a downward pitch contour. However, what we see in the word 什么 is that the syllable *me* had a high pitch target in all the phonetic contexts.

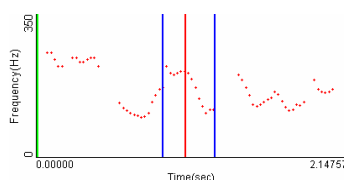
One may argue that the syllable *shen* was unstressed or in neutral tone status in those sentences, that resulted in a low pitch contour in *shen*. However, perception of the sentences did not show that the syllable in those sentences was unstressed at both syllable and word levels. On the other hand, the syllable *shen* could not be in neutral tone in that particular word. Because that is against the basic assumption for a *Qingsheng syllable*, which normally only happens in word middle or final position.

Another possible argument could be that it is a tone co-articulation phenomenon. Tone2 could have a level or even a little falling pitch contour when it is in some particular phonetic context. If the low pitch in *shen* was caused by tone co-articulation, then it should be either affected by the pitch in its preceding or its following syllable. However, as shown in figure 2, the syllable *shen* kept roughly the same low pitch contour when it was preceded by different lexical tones. Graph [a] in Figure 2 is a null preceding tone case, that is, the syllable *shen* was at the very beginning of a sentence. In that situation, there could be no preceding tone effect. Therefore, the possibility of tone co-articulation by its preceding pitch does not stand. The co-articulation effect from the following tone could also be dismissed because the following syllable *me* was in neutral tone. According to neutral tone rule, *me* has no pitch pattern of its own, but follows the pitch pattern of its preceding syllable *shen*. Consequently, the lexical tone pattern in *shen* could not be affected by the syllable *me*.

An alternative interpretation, then, is that there possibly exists unidentified tone *sandhi* patterns in modern spoken MC, which were neglected in previous linguistic studies. The reason could be that those tone sandhi patterns only appear in a few specific words and from linguistic point of view, those tone *sandhi* may not be phonologically distinctive. If this interpretation stands, then, there could be two possible hypotheses for this tone sandhi case. One is that this tone sandhi pattern applies to any Tone2 + neutral tone word. The other is that this tonal pattern is unique which only exists in some particular words like *Shenme*.

For the first hypothesis, we examined *Qingsheng* words with the same Tone2 + neutral tone pattern in the real speech data. Figure 3 illustrates two such cases.

[a] *Qingsheng* word 答应 (Daying) in the sentence
对方爽然答应派人来津



[b] *Qingsheng* word 鼻子 (Bizi) in the sentence 在场的不少人鼻子酸酸的。

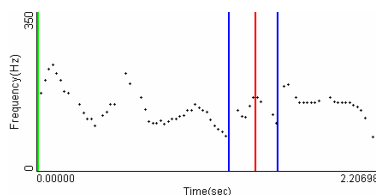


Figure 3. Tonal patterns of the *Qingsheng* words containing Tone 2 syllable in word initial.

Figure 3 shows that though *Tone2* pitch pattern in the word 答应 as shown in the graph [a] was somewhat flat due to its tone coarticulation with the preceding *Tone2*, the following neutral tone was still kept falling, rather than a raised pitch in the *Qingsheng* syllable as in the word 什么. The same was for the neutral tone pattern in the word 鼻子 as shown in the graph [b], where the neutral tone had a falling pitch. Therefore, the tonal pattern in the word 什么 could not be a universal tone sandhi phenomenon. Then, the possible *Tone sandhi* interpretation is the second hypothesis which states that the tonal pattern in word 什么

is a unique tone sandhi case only existing in some particular words.

From the practical consideration of speech synthesis, we currently in our parameter-based TTS system defined a new low level tone pattern and a corresponding tone *sandhi* rule which specified that if the neutral tone is preceded by that low tone, the pitch should be raised above the endpoint pitch of the preceding syllable to certain degree. This is similar to the neutral tone rule for the syllable preceded by *Tone3*. With this strategy, the synthesized word 什么 sounded quite natural, and the TTS system is also capable to handle new word cases similar to 什么 in future. However, the linguistic interpretation of this lexical tone exception phenomenon still needs further investigation and discussion.

4. Conclusion

In this paper we have presented and discussed two MC tonal issues which occurred in our MC TTS synthesis experiment. There are other lexical tone related speech synthesis issues, such as the tonal patterns in different affective states, which are currently under investigation at our research lab.. We hope our TTS research work on the tonal aspect of MC will benefit our understanding of the MC spoken language, as well as improving the naturalness of the synthesized speech.

5. Reference

- [1]. Jie Zhao, 2001, Chinese Linguistics, Zhao Hua Publishing House.
- [2]. Cao, Jianfen 1992. On Neutral-Tone Syllables in Mandarin Chinese. Canadian Acoustics, Vol. 20, No.3.
- [3]. F. Chen, 2002, Issues in Speech Synthesis for Tonal languages. SNLP-Oriental COCOSDA 2002, Thailand. p. B1-6.