The Duration Analysis of the Checked Tones in Cantonese Speech

Xiaoying Xu (1) Jianhua Tao (2) Ling Zhang (3) Yingchao Lu (4)

(1)(3)(4) PSC Center, Chinese Language & Literature Dept.
Beijing Normal University, 100875, Beijing
(1)xuxiaoying2000@bnu.edu.cn
(3)d_r_036@yahoo.com.cn
(2) jhtao@nlpr.ia.ac.cn

Abstract—it is well known that the duration reflects the difference between checked tone’s syllables and other syllables in the Cantonese TTS system, we use two concepts, the voice blank and pitch contour blank, to examine the particularity of the checked tone’s duration with a serial designed experiments in Cantonese speech. The impact of the four factors, including the tone types of the Cantonese checked tone’s syllable (CHKTS), the types of the occlusive stop of the CHKTS, the onset types of the syllables following the CHKTS, the location in the sentence of the CHKTS, have been investigated for the duration of check tones. The results are helpful to construct the context information for the prosody prediction in Cantonese TTS system. Moreover, with two perception experiments, the paper also proves that the silence segment which indicated as voice blank and extension of the onset duration following the CHKTSs in the Cantonese utterance is a kind of prosodic compensation of the CHKTSs.

Keywords—Cantonese; Checked tone words; Voice blank; Duration; Pause

I. INTRODUCTION

The tones of Cantonese can be divided into two types: (i) the checked tone, namely syllables ending in an occlusive coda (-p,-t,-k); (ii) the smooth tone, namely either an open syllable, or a syllable closed by a nasal stop. It is well known that the checked tone syllables (CHKTSs) are obviously shorter than other tone syllables when they are read isolated [13]. In most case, there is 0.1 second difference between these two tone types. This difference can be obviously perceived by human. Thus, the duration plays an important role in differentiating their lexical meanings, and has the strong impact for the prosody processing of checked tones in Cantonese TTS system.

There are large amount of documents focusing on the duration of isolated syllables and continuous speech in the tonal language (e.g. [2][3][4][7][8][9][15][16]). In their work, they made the analysis of syllabic duration patterns [2][4][8] in Mandarin speech, and tried to construct the model for the duration prediction for Mandarin TTS [3][7][9][15]. The results have greatly improved the naturalness of Mandarin TTS systems. However, there is still little work focusing on the duration analysis of Cantonese continuous speech [13].

In the paper, aiming at improving the naturalness of the Cantonese TTS system, we use two concepts, the voice blank and pitch contour blank, to examine the particularity of the checked tone’s duration. Voice blank, which appears as a non-wave period between two syllables in the spectrum, means the silent section between the two syllables. Pitch contour blank, which appears as the distance between two adjacent syllables pitch contour, means the non-voice section between two syllables. Via serial experiments, the paper finds,

(a) Comparing with the smooth syllable, the checked tone’s syllable follows with either a larger voice blank or a larger duration of the onset in the following syllable. We consider these two phenomena as the particularity of Cantonese checked tone’s syllables (CHKTSs) in continuous speech.

(b) Four factors, including the tone types of the Cantonese checked tone’s syllable (CHKTS), the types of the occlusive stop of the CHKTS, the onset types of the syllables following the CHKTS, the location in the sentence of the CHKTS, have been investigated for the duration of check tones. The results show that there is no obvious impact for the tone type on the duration of checked tones. Nevertheless, the types of the occlusive coda are more important. Among all features, the onset types of the syllables following the checked tone syllable (CHKTS) have most impact on the particularity of the checked tone’s duration. If the CHKTS is in the position of the right boundary of the prosodic units, it will obey the general rhythm principle of the right prosodic boundary.

(c) The phenomena of voice blank and the extension of the onset duration following the checked tone can not be attributed to the function of syntax, logic, semantics or stress, but rather to the function of the sentence rhythm. It can be thought as a kind of duration compensation of CHKTSs, and plays a role that helps the CHKTS to consist a normal foot in the sentence.

The rest of the paper is organized as follows: Section II examines the factors which affect the duration of the checked tones; Section III discusses the rhythm characteristics of the CHKTSs’ duration. The final conclusion is made in Section IV.

II. THE FACTORS AFFECTING THE DURATION OF THE CHKTS

As the four factors, including the tone types, the types of the occlusive stop of the current syllable, the onset types of the
following syllable, the location of the syllable in the word and sentence, are normally important to form the context information to predict the prosody parameters in TTS. We easily think about how about the impacts of these factors are for the duration distribution of Cantonese checked tones. To make the analysis, we designed a serial Cantonese sentences and each sentence was recorded by two professional Cantonese speakers (one male and one female).

A. The tonal types of the checked tones

As we know, checked tones of Cantonese can be divided into three types according to the tone degrees of pitch value (normally labeled as 5, 3, and 2) in the syllables. Besides these three types, another checked tone of with the degrees 35 pitch value will occur in the real stream appear under the tone Sandhi processrules. Thus, there are a total of four checked tones types in Cantonese. In this part, the designed sentences used for the analysis are listed as followings. In this experiment, informants were asked to read the following sentence:

a) 透过[k5] to33 kai55 uk5 kei35
   (The measurement work has been started at his home.)

b) 透过[k5] to33 kai55 uk5 kei35
   (The removal work has been started at his home.)

c) 透过[k5] to33 kai55 uk5 kei35
   贼 (平) 到 佢 屋 企
   (The thief enters his home.)

d) 透过[k5] to33 kai55 uk5 kei35
   贼 (升) 到 佢 屋 企
   (The thief enters his home.)

Among these sentences, all second syllables, “to33”, which have both the same segmental structure and the same tone degrees, are the smooth tone, and all of the first syllables, “kai55”, “kai3”, “kai2” and “kai35”, are the checked tones. With the speech, recorded by a male and a female (as mentioned above), we then get all the voice blanks following the check tones, and the pitch contour blanks. The results are shown in Table I.

<table>
<thead>
<tr>
<th>Word</th>
<th>Voice blank following the checked tones (s)</th>
<th>Pitch contour blank (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ha53] (bully)</td>
<td>0.144 0.103 0.168 0.100</td>
<td></td>
</tr>
<tr>
<td>[ha5] (black)</td>
<td>0.147 0.102 0.168 0.110</td>
<td></td>
</tr>
<tr>
<td>[ha35] (beg)</td>
<td>0.157 0.103 0.170 0.110</td>
<td></td>
</tr>
<tr>
<td>[hap5] (bully)</td>
<td>0.126 0.087 0.159 0.101</td>
<td></td>
</tr>
</tbody>
</table>

From Table I, it shows that both the voice blanks following the checked tones and the pitch contour blanks are similar, except the voice blank following the CHKTS “贼 (rising)[shak35]” is a little bit less than the others. In general, we can think that there is no obvious impact between the tone type and the duration of the syllables with checked tone.

B. The coda types of the CHKTS

In this part, we try to design an experiment to investigate the relative relation between the types of the occlusive stop coda and the duration of the CHKTSs. For the occlusive stop coda in Cantonese, there are totally three types, [-k],[-t] and [-p]. The sentences used for analysis are:

a) 透过[k5] tou33 kai35 haam33
   (He is bullied to cry.)

b) 透过[k5] tou33 kai35 haam33
   (He is scared to cry by the dark sky.)

c) 透过[k5] tou33 kai35 haam33
   (He cries with the tired begging.)

d) 透过[k5] tou33 kai35 haam33
   (He is bullied to cry.)

In these utterances, except the first syllable of sentence “a” is the smooth tone’s syllable, all other first syllables among these sentences (from sentence “b” to sentence “d”) represent the CHKTSs with different types of occlusive stop coda. With these samples, the results of the voice blanks and the pitch contour blanks are got in the following table.

<table>
<thead>
<tr>
<th>word</th>
<th>Voice blank following the checked tones (s)</th>
<th>Pitch contour blank (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ha53] (bully)</td>
<td>0.047 0.026 0.078 0.031</td>
<td></td>
</tr>
<tr>
<td>[ha5] (black)</td>
<td>0.172 0.109 0.209 0.141</td>
<td></td>
</tr>
<tr>
<td>[ha35] (beg)</td>
<td>0.219 0.150 0.248 0.169</td>
<td></td>
</tr>
<tr>
<td>[hap5] (bully)</td>
<td>0.182 0.136 0.209 0.159</td>
<td></td>
</tr>
</tbody>
</table>

From Table II, it shows that the voice blanks following the CHKTSs are much larger than the voice blank following the smooth tone’s syllables, and the pitch contour blanks have the similar conclusion as the voice blank. We can then think that the types of the occlusive coda have relatively more obvious impact comparing with the factor of tone types.

C. The onset types of the following syllable

To analyze the impact of the onset types for the duration of CHKTSs, we selected a typical syllable, “— iat5 (one)” which contains the checked tone because it can collocate large amounts of measure words and can be easily put into the sentences.
The structure of the designed sentences is as follows:

a)  

\[ \text{sei 33 fan55 ts5i55 i4t5 ko33 dan11 kou55} \]

(He/She) just bought a piece of cake.

b)  

\[ \text{ngam55 mai5 ge33 iat5 ko33 dan11 kou55} \]

岩买了一个蛋糕

c)  

\[ \text{Wai21 i4t5 kuaan55 sam55 ge33 tsai33 hai22 kai23} \]

唯一关心既就是

He is the only person caring about it.

d)  

\[ \text{ni55 iat5 kua:n55 m21 hai22 kam33 i33 kuo33} \]

唯一关塔系唔易过

(That barrier is not so easy to be passed.)

e)  

\[ \text{kai23 sing55 kei21 iat5 dit3 tsan33 tso35 koek3.} \]

星期一跌亲左脚

(He broke his left foot on Monday.)

f)  

\[ \text{kai23 ge33 ni55 iat5 dit3 jau23 paai21 i55} \]

既呢一跌有排医

(It takes him a long time to cure for his broken leg.)

Again, we use the numeral “- [iat5] (one)” to construct a serial sample sentences. Each sentence contains a different prosody boundary structure where the voice blank following the checked tone is located in. For instance, in Sentence “a”, the CHKTS “- [iat5] (one)” and the following syllable belong to the different prosodic words. This sentence represents the structure that the voice blank is located at the boundary of inter-prosodic-word. In Sentence “b”, the CHKTS “- [iat5] (one)” and the following syllable belong to the same prosodic word, thus the sentence represents the structure that the voice blank is at the boundary of intra-prosodic-word.

In this part, the prosody structures, where the voice blanks located in, of all sentences are classified into three types, intra-prosody-word, inter-prosody-word, inter-prosody-phrase. The inter-prosody-phrase means the voice blank of the checked tone is located in. For instance, in Sentence “a”, the CHKTS “- [iat5] (one)” and the following syllable belong to the different prosodic words. This sentence represents the structure that the voice blank is at the boundary of intra-prosodic-word.

With these different prosody structures among all sentences, we get the following the results.

### Table III. The Results of the Onset Types of the Syllable Following the CHKTS

<table>
<thead>
<tr>
<th>Types of the onset</th>
<th>Duration of onset of the following syllable (s)</th>
<th>Voice blank + Duration of onset of the following syllable (s)</th>
<th>Pitch contour blank (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unspirited stop</td>
<td>0.007~0.03</td>
<td>0.07~0.1</td>
<td>0.077~0.13</td>
</tr>
<tr>
<td>Aspirated stop</td>
<td>0.04~0.07</td>
<td>0.03~0.06</td>
<td>0.07~0.13</td>
</tr>
<tr>
<td>Affricate stop</td>
<td>0.04~0.07</td>
<td>0.03~0.07</td>
<td>0.07~0.14</td>
</tr>
<tr>
<td>Fricative</td>
<td>0.09~0.11</td>
<td>0.09~0.11</td>
<td>0.09~0.11</td>
</tr>
<tr>
<td>Nasal and glide</td>
<td>0.06~0.13</td>
<td>0.04~0.05</td>
<td>0.1~0.18</td>
</tr>
<tr>
<td>Semi-vowel</td>
<td>0.05~0.1</td>
<td>0.06</td>
<td>0.11~0.16</td>
</tr>
</tbody>
</table>

From Table III, we can conclude that: (i) In normal speed, the variation of the “Voice blank following the CHKTS + Duration of onset of the following syllable” is relatively consistent, most of them are within a narrow range, 0.07 to 0.14 second. (ii) Except the semi-vowel, the “Pitch contour blanks” are also relatively consistent, which is nearly 0.1 second. (iii) For the consonants, “Unspirited stops”, whose durations are short and their duration can rarely be changed in the stream, they have a relatively long voice blank following the CHKTSs; For the consonants, “Fricative”, whose duration are comparatively long, it’s even not necessarily to have a piece of voice blank.

From the above, we can see the onset types of the syllables following the CHKTS have the strong impact on the particularity of the CHKTS duration.

### D. The prosody boundary types

1) The boundary of different prosody structures

To analyze the impact of the boundaries of different prosody structures for the duration of CHKTSs, we design the sentences as follows,

a)  

\[ \text{sei 33 fan55 ts5i55 i4t5 ko33 dan11 kou55} \]

四分之一个蛋糕

(A quarter of the cake.)
From Table IV, comparing with Sentence “b”, Sentence “a” contains a larger “Voice blank + Duration of onset”, which indicates that the voice blank following the checked tone can also be used as the silence of prosody boundaries. While comparing among the three types of prosody structures, we can find the “Voice blank + Duration of onset” is longest in inter-prosody-phrase structure. It follows the normal rules of silence duration of the prosody structure. However, we found there is no obvious difference of the “Voice blank + Duration of onset” between the inter-prosody-word structure and the intra-prosody-word structure in male speech, while the “Voice blank + Duration of onset” of the prosody-phrase structure is obvious longer than that of the intra-prosody-word structure in female speech. The reason is that the speed of the male speech is much faster than the female speech. It conceals the duration differences in the male speech.

The results indicate that if the CHKTS is at the boundary of the prosodic units in normal speed, it normally obeys the general duration distribution of the rhythm principle.

2) The position of before ward and afterward in the sentence

In this part, we try to inspect the different impact between the position of before ward and afterward in the sentence. We design seven sentences, the position of prosodic word “一部[iat5 bou22] (one)” which contains with the CHKTS “―[iat5](one)” are arranged to move step by step from the head to the final of the sentence gradually.

The sentences and the CHKTS’s positions are as follows:

a) iat5 pou22 sau23 ki55 tou55 maa22 m21 tsa22.

(No hand phone can be sold out.)

b) Mui23 jat6 maai22 tsa22 iat5 pou22.

(One (hand phone) is sold out per day.)

c) Mui23 jat6 maai22 tsa22 iat5 pou22 syn23.

(One book is sold out per day.)

d) Mui23 jat6 maai22 tsa22 iat5 pou22 sau23 ki55.

(One hand phone is sold out per day.)

e) Mui23 jat6 maai22 tsa22 iat5 pou22 tin22 peng55 saeng55.

(One refrigerator is sold out per day.)

f) Mui23 jat6 maai22 tsa22 iat5 pou22 meng21 pai21 sau23 ki55.

(One brandy hand phone is sold out per day.)

g) Mui23 jat6 maai22 tsa22 iat5 pou22 meng21 pai21 tin22 peng55 saeng55.

The result is shown in Table V.

<table>
<thead>
<tr>
<th>Number of sentence</th>
<th>Voice blank (s)</th>
<th>Duration of “―[iat5]”+voice blank (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>a</td>
<td>0.126</td>
<td>0.134</td>
</tr>
<tr>
<td>b</td>
<td>0.130</td>
<td>0.140</td>
</tr>
<tr>
<td>c</td>
<td>0.103</td>
<td>0.107</td>
</tr>
<tr>
<td>d</td>
<td>0.105</td>
<td>0.115</td>
</tr>
<tr>
<td>e</td>
<td>0.088</td>
<td>0.099</td>
</tr>
<tr>
<td>f</td>
<td>0.083</td>
<td>0.064</td>
</tr>
<tr>
<td>g</td>
<td>0.070</td>
<td>0.077</td>
</tr>
</tbody>
</table>

In the first and second sentences, “―[iat5 pou22](one)” is in the head and final of the sentence respectively, and their “duration of “―[iat5]”+voice blank” are relatively larger than other sentence; In sentences “c”, “d”, “e”, “―[iat5 bou22](one)”is in the afterward position of the sentences, their value of the “duration of “―[iat5]”+voice blank” are relatively small; In sentence “f”, “g”, the value is smallest.

According to Cao’s point [2], the syllable’s duration will be obviously decreased when in the middle of the sentence while its duration will be obviously increased either in the head or in the final of the sentence. Our data are consistent with this inspect, the duration of the CHKTS and the syllables following it also changes according to the rhythm of the sentence in Cantonese.

III. THE RHYTHM CHARACTERISTICS OF THE CHKTS’ DURATION PARTICULARITY

The above experiments have proved that voice blank is important for the duration analysis of the checked tones. The question is if the voice blank can be simply thought as a pause. In this section, we try to answer this question via two perception experiments. The testing sentences used here include both the sentences used in the above sections and some more sentences chosen from NLPR Cantonese Speech Corpus. In each testing sentence, no CHKTS is in the prosodic boundary, and the voice blanks following the CHKTSs are longer than 0.1 seconds which can be obviously perceived in ordinary context. The 16 informants are asked to judge whether there is a pause and label the location of the pause in the sentences. In order to exclude the impact based on the understanding of the meanings and the syntax structure, we choose someone who does not understand Cantonese as the qualified informants to ensure that their judgments are only made based on the speech rhythm aspects. The results show that although the informants don't understand Cantonese, they can judge the semantics and syntax pause correctly while none of them recognize the voice blank following CHKTS which isn't at the prosodic boundary as a pause.
As concerning the conceptual categories of pause, many researchers have only treated it as a kind of acoustic phenomenon, and most of them have inspected it from the viewpoint of "sound blank." But Ye [14] pointed out that pause is not a acoustic concept but a perceptual concept, and the acoustic correlation factors which form the pause perception include not only the utterance interval." That means the essence of pause is a perceptual concept, though its objective and external features are reflected to the acoustic categories. In other word, pause is the result of acoustic categories projecting onto the perceptual categories. Researchers can use acoustic parameters to measure and study pause in most of the circumstance, but in the speech communication process, the realization of pause's function depends on the listener's mental cognition. Summarizing the views of the researchers, we can see that there are mainly three kinds of performances of pause in the utterance: (1) silence segments; (2) the extension of following syllable; (3) the voice register contraction of the whole rhythm group [6][11][14]. As concerning the functions of the pause, there are also three kinds of functions: (1) syntactical (logical); (2) semantical (emphasize); (3) physiological. [6][11][14].

Let's turn back to the analysis on the voice blank following the CHKTS. Firstly, view from acoustic categories, it is indeed shown as silence segment, the voice blank following the CHKTS of the testing sentences is longer than 0.1 seconds. But in the perceptual categories, the perception experiment in this section proved that the voice blank can be hardly perceived as a pause. Due that the determinant factor for a phenomenon to be considered as pause is to be perceived by human beings, the voice blank following the CHKTS can not be treated as a pause. Secondly, from the functional point of view, the voice blank following CHKTS does not only appear in the prosodic boundary or the position where needs to be emphasized, so this voice blank can not be attributed to the needs of syntax, logic, semantics or stress. In addition, when people talk they usually take a breath after more than ten to twenty syllables, but in our testing sentences, some phrases only consists of checked tone words, while some other sentences consists of several syllables, so the voice blank in these kind of sentences can't be interpreted physiologically. In a word, the voice blank following CHKTS is not a pause, at least it is not such kind of pause which can segment the grammatical structures or emphasis the semantics as researchers pointed out broadly.

Ye [14] thought "silence segments are not always the mark of separating the utterance stream. For instance, the silence segment in front of the plosives stop only indicates that the VOT value is 0 or negative, and it does not have the function to separate the lexical units." Zhou [17] pointed out that " Plosives or affricates of consonants in removing resistance period often results in short intervals, which shown in the spectrum as vertical blank space on the time axis." At the end of CHKTS, there is a consonant stop coda, which only has a very short period of forming the resistance and supporting the resistance but without the removing resistance period, which means it doesn’t pronounce, so the only function of the voice blank following the CHKTS represents the period of forming the resistance and supporting the resistance of the consonant stop coda of the CHKTS. However, this can only account for unaspirated stop onset following the CHKTS and can’t account for other types of consonant onset and semi-vowel onset. For the following fricative, nasal, lateral and semi-vowel onset, the voice blank are near zero, but the duration of the following onset has been extended obviously.

Therefore we can only seek the reason from the rhythm of the sentence. According to Feng’s point [5], the emergence of the tone makes the whole syllables under the tone’s government. Thus the syllable’s length is no longer calculated by the number of internal mora, and the mora can not indicate the length differences of syllables, so there is no weight difference in the interior of syllables, it is impossible to construct a mora foot in interior of the syllables. As a result, syllable foot will be constructed instead. The length of four tones in Mandarin have few differences in duration which can be neglected by perception. In this sense, we can say that whatever the tonal types are, syllables in Mandarin are uniform structures with equal status and have no weight differences. Only in specific contexts there will be weight differences and then forming the foots, but the formation of foots has nothing to do with the tonal types.

The prosodic structure in Cantonese is also the binary foot structure, but the tonal condition is different from Mandarin. As we analyzed above, the duration of CHKTS differs a lot from other tonal type syllables, which can be perceived by listeners. So in the viewpoint of syllables’ duration, it is difficult to consider all the tonal syllables into uniform structures. If the syllables are spelt together directly, the status of checked tone words and that of other tonal types is unequal, the CHKTS’ duration is shorter and lighter than other tonal types, that is to say, they can form foots by tonal differences in specific environment, which is not consistent with the rhythm of Mandarin.

We can get the direct understanding about the analysis from Figure 1. In order to get equality with other tonal types, the tone contour blank between the checked tone words and the following syllables has to be extended, and then they can add the duration of tone contour blank to their own governmental area so that they can get the similar duration with other tonal type words and get an equal status. There are two ways to achieve this goal: one is to increase the voice blank; the other is to extend the duration of the onset of the following syllable, because it isn’t the main undertaker of the tone.

We continue to perform another perceptual experiment to test our conclusion. Three types of checked tone materials are used in this experiment: (i) The isolated CHKTS; (ii)The sentences recorded for the above-mentioned experiments, for
instance, “黑到就喊”，“治到就喊”等, etc, we call these sentences as the nature rhythm sentence; (ii) Cut the voice blank following the CHKTSs of the nature rhythm sentences and forming a set of new sentences. Ten native Cantonese informants are asked to repeat the sentences and judge whether the sentences are normal rhythm sentences or abnormal one.

As the experimental results shown, most of the subjects can distinguish the abnormal rhythm when listening to the sentences without voice blank. Both the isolated CHKTS and the CHKTS in nature rhythm sentences can be repeated correctly. But most of informants can’t distinguish the CHKTSs in the sentences without voice blank, about half of them can’t distinguish “治” and “乞” from “哈”, about 81% confused “黑” and “哈”. That is to say they confuse the CHKTSs with the corresponding smooth tone syllables.

We can explain the experimental results according to the theories referred above. The pitch contour blank controlled by the CHKTS is an important feature for distinguishing the CHKTS. When listening to an isolated CHKTS, there will be a silence for the listener to make judgment, but in the sentences without voice blank, the CHKTSs are connected directly with the following syllable and will be perceived as a disyllable words.

The analysis in this section can also be used to analyze an progressing trend accrued in Cantonense that the checked tone changing to the smooth tone. As Yang [13] documented, there are some kinds of changing trends of the checked tone, one of them is Canton-style trend, which is “the occlusive coda of the checked tone disappeared, while the checked tone preserved its original pitch value and merge into the corresponding smooth tone type.”

According to the experiments and analysis in this article, we can see that the so called Canton-style trend is the materialization of the voice blank following the CHKTSs in fact, it eliminates the duration differences between the different tone types and makes all the tone types into a generally similar duration construct, finally it forms a normal syllable foot, which is consistent with the development trend of rhythm.

IV. Conclusion

Via a series experiments, this paper finds: Firstly, the particularity of CHKTSs’ duration in utterance include two phenomena which are following with a larger voice blank and the duration of the onset of the following syllable will be extended. Secondly, among the fours factors we investigated in this paper, the tone type, the types of the occlusive coda have relatively little impact on the duration particularity of the syllables with checked tone, while the onset types of the syllables following the CHKTS and the position of CHKTS have great impact on the particularity of the checked tone duration. Finally, the phenomena of voice blank and the extension of the onset duration following the checked tone are attributed to the function of sentence rhythm, it’s a duration compensation of the CHKTS, and plays a role that helps the CHKTS to consist a normal foot in the sentence. The results are helpful to construct the context information for the prosody prediction in Cantonese TTS system.

The results of this paper can also overcome some rhythm defaults found in a Cantonese TTS System which include: (i) The speed of some synthetic sentences are too fast, which is particularly obvious in technical terms and infrequent words. (ii) As concerning as the onset categories of the syllable following the unrhymed checked tone words, though unaspirated plosives [p], [t], [k] are more common, it can also occur in aspirated plosives, affricates, fricatives, nasals, laterals, and semi-vowel etc. (iii) In general, the synthesis of common usage words and sentences is better than infrequent words and sentences.

Acknowledgment

The work was supported by the National Science Foundation of China (No. 60873160, 90820303), 863 Programs (No. 2009AA01Z320).

References