RPR—IL(CASS)/1990,20—34

THE DYNAMIC MODEL OF SYLLABLE DURATION IN STANDARD CHINESE

Cao, Jianfen

汉语普通话音节时长变化的动态模式

曹剑芬

[详细提要] 在中国，关于语音协同发音的研究方兴未艾，而且，现有的研究多半注重于它的空间方面。事实上，语音也同宇宙间的万事万物一样，处于不断的变化发展之中。这种变化和发展，归结起来不外乎空间的和时间的两个领域。因此，对于语音协同发音的研究也离不开空间域和时间域两个方面。而且，这两方面的特点是密切相关、相互依存的，离开了时间域的特点，无以观察空间域特性的动态变化；同样，离开空间域特性的变化，时间域的特征也无从计量和表达。所以，它们之间的关系尤如一张纸的两个面，相辅相成，谁也离不开谁。正因为如此，有必要对语音的时间域特性或者说时间维特性进行更加深入的研究。

广义地说，语音的时间维特性研究应当包括两个大的方面：一个方面是研究语流中语音的各项空间特性的定时关系，即关于这些特性协调变化的时间控制，诸如语音的音色、音高和音强等特征要素动态变化的时间分布，以及相邻语音单元的这些特性在时间轴上的相互影响和叠接关系；另一个大的方面是研究各语音单元之间的协同调时效应，即研究语音的时长特征本身在语流环境中的协调变化，诸如音段、音节、词等语音单元在不同语境和不同条件下的时长变化及其规律。本文的研究基本上属于后者，这里将集中讨论汉语普通话里音节时长变化的动态模式及其相关的控制因素。

在自然语言里，音节的时长变化具有极大的动态范围。就汉语普通话而言，根据我们的初步观察，用中等速度独立单念的音节多半在 500—600 毫秒上下；而在自然的连续话语里，作为重读的单音节词的音节多半在 300—400 毫秒之间，而那些非重读的音节则要短得多，有的甚至不到 100 毫秒长。然而，尽管如此，语言仍然具有一定的节律，而且，它们的速率通常也保持在一定的范围以内。那么，在自然语言里，音节的时长究竟是怎样变化的？遵循甚么样的规律？是哪些因素造成了如此悬殊的时长差异？它们是怎样影响音节的时长变化、又是怎样控制语言的节律的？为了对这些问题有个基本的认识，我们通过对一个女发音人和两个男发音人所发的十四对不同类型的三音节词和六个句子的实验分析，并参考我们先前的相关研究，对汉语普通话的音节时长在不同语境和条件下的具体变化及其规律以及影响这些变化的控制因素及其影响方式作了初步考察。

根据本实验所获得的结果，对普通话音节时长的动态变化有如下几点认识：

第一，平均时长：在实际语言里，由于音节的时长具有极大的伸缩性，因而很难给出确切的平均时长来。不过，根据我们的实验资料，仍然有个大致的常数：在中速话语里，每秒钟说出的音节数目在5—6个之间；在快速话语里，每秒钟可发6—7个音节。根据这个情况，并考虑到音节之间发音上的相互叠接，若将这个部分也分别计入各自的音长，那么，普通话语流中音节的平均长度大约为200—250毫秒。

在一般情况下，这个平均时长似乎并没有甚么实际意义。不过，每秒钟大约能说5—7音节，这样一个平均速率或许并不是偶然的。言语产生和言语感知的相关研究发现，幼儿呀呀学语时每次连续发出的音节数从不超过7个；而人的整块记忆的跨度通常也是包含7个左右音节。所以，人们猜测，这个数目可能表明，在人类听觉理解的过程中，存在着同携带单位意义的语法单元相应的知觉单元，而这个单元一般包含7个左右音节，这可能是在时域控制方面存在于各种语言里的一个共同基础。我们都知道，人类言语的产生同感知是密切相关的，因此，它们应当具有共同的时间控制机制。所以，我们可以推测，这里观察到的言语平均速率可能反映了自然语言时域控制方面存在的一个相对不变量。这样，与此相应的音节平均时长或许可以作为语流中衡量音节时长伸缩变化的大致标尺。

第二，固有差异：对于不同语言的许多研究发现，不同类型的音节之间存在着固有的时长差异，它们是由音节本身的语音特性或音系特点之类的因素决定的。在这些因素中，语音性质方面的影响，诸如辅音的发音方法或元音的口开度等，它们的作用是跨语言的，普通话里也是如此。例如，一个具有送气塞擦音声母和鼻尾韵母的音节，假如其它条件相同，那么，它往往会比别的类型的音节明显地长些。相对说来，音系结构方面的影响往往是因语言而异的。对于汉语普通话来说，声调的区别和轻重音的对立是另外两个不可忽视的重要因素。一般说来，上声音节普遍比其它音节要长，而去声音节则明显地短些；同样，轻重音的对立也直接影响音节的固有时长差异。我们的有关分析表明，在普通话里，正常重音音节同轻声音节的时长比大约为100:63(单念情况下)到100:50(多半在语流中)。

第三，音节在语流中的时长变体：首先，孤立单念的跟语流中的音节之间存在着系统的差异。以单音节词为例，中速单念的音节对连读的音节，其平均时长之比大约为100:70；其次，语流中的音节时长变化主要受语境条件的控制，诸如音节在语流中的位置的不同，语速的变化和语音单元的大小，语音单元的结构特点，以及其它语法和语义等语言学因素的作用。位置、语速和语音单元大小对音节时长的影响是不同语言里所共有的现象，而语言学因素的作用则取决于不同语言的具体情况。

(1) 位置效应：通常，词末、短语末或句末的音节比相应的起首位置上的要长些，而处于中间位置上的音节常常比一般的要短些。比较表I里的各行数据，就可以清楚地看出，普通话里音节的时长差异同样显示出这种效应。但是，这种影响会受到某些语言学因素的制约。关于这一点，将在第(3)部分详细说明。

(2) 语音单元的长度和语速的影响：一般说来，音节的时长随着语音单元的加长而减缩，也就是说，它是随着单元内音节数目的增加而缩短。表I下部(a)(b)两行数据之间的差异就说明了这种倾向；此外，音节的长度会随着语速的加快而变短，比较表I左右两栏里的数据，这种系统差异是显而易见的。然而，通常人们感觉到的语速变化，往往会通过句中短语之间可能出现的、或长或短的停顿来调节。图2是用两种不同速度所说的同一句话的语图，比较这两种情况下句子总长之差和停顿的差异，我们不难看出，音节长度的伸缩同语速的快慢之间的反比关系并非完全线性的。

(3) 语音结构特点的制约：

在语流中，音节多半作为更大语音单元的一个部分活动，它的时长往往取决于那些单元的结构特点，因而相应地构成了比较稳定的时长分布模式。图 1 就是普通话双音节词和三音节词里音节时长模式的图示。值得注意的是，在汉语普通话里，双音节词是个相对稳定、而又具有极强构词能力的结构单元，它同单音节词一起生成其它的多音节结构。因此，作为基本的构词单元之一，它的时长模式也具有相对的稳定性。所以，较大的语音单元的时长分布模式常常同双音节词的模式密切相关。例如，在图 1 里，A 和 B 分别代表正常重音型和轻声型双音节词的时长模式，其余的是由单音节词和双音节词通过不同的方式生成的三音节词的时长模式，从这些不同的三音节时长结构中、尤其是带轻声音节的结构中，可以明显地看出上述倾向。譬如，B(1)中的第二音节和 B(2)及 B(3)中的第一音节并没有因为处于词首或词的中间位置而缩短；相反，B(1)和 B(3)中的第三音节也并没有如预期的那样明显延长，尽管它们都处于词末的位置上。这些现象说明，位置对于音节时长的影响还是有限的，它还要受到语音的结构规则的制约。

(4) 句法环境的制约：在语流中，音节的时长还直接受其后随句法成分特性的影响。例如，我们作了一个比较，当一些音节后随的是个独立结构时，这些音节的平均时长是 292 毫秒；而当后随的是个语尾助词时，它们的平均长度缩短到 254 毫秒。这是因为，语尾助词一般不能独立，它们必须附着在前面的结构上作为这个结构的一部分，这样，就使前临音节处于词中间的位置上，因而使它的时长受到限制。

(5) 底层语义要求的控制：语音是底层语义内容的表层表达，因此，自然语言里语音的表层实现必然同底层的语义要求密切相关，音节时长的具体实现当然也不例外。从我们的实验结果来看，底层语义结构对于音节时长的控制通常是通过不同的语句重音来实现的，具体情况如图 3 所示。在这里，有四个语义内容不同的陈述句，它们都是由同样的音节系列 '这 双 鞋 不 结 实' 构成的。第一句话的基本语义是一般性质描写，回答 '这双鞋怎么样？' 的问题，语句重音落在谓语部分，即落在 '不结实' 这个短语上。而在这里关键词是 '结实'，它是个轻声词，根据词重音规则，'结' 字重读，而 '实' 字读轻声。因此，这个句子的强调重音就落在了 '结' 这个音节上；第二句话是回答 '到底结实不结实?' 的问题，这里强调 '不' 结实；第三句针对 '甚么东西不结实?' 的问题，强调这双 '鞋' 不结实；而第四句则是回答 '哪双鞋不结实?'，这里强调 '这' 双鞋不结实。从图 3 我们可以看到，由于这四句话的底层语义要求的区别，通过短语、词和音节各平面上重音规则的作用，最后形成了不同的句子重音分布，各句的强调重音分别落到了音节 '结'、'不'、'鞋' 和 '这' 上，并由此构成了音节时长的不同表层实现。从这个例子还可以看出，在影响音节时长的众多因素中，语义要求是个最为活跃的因素，其它各种因素都在它的控制下发挥作用。

综上所述，在汉语普通话里，影响音节的时长变化的主要因素可以归纳为内在的和外部的两个方面，内在的是同音节本身相关的语音学的和音系学的因素，外部的主要是语言学的因素。外部影响通过内在因素起作用，而内在因素又必须在外部影响的控制下发生效应。也就是说，所有这些因素的作用一方面是分层次、分等级的，它们各自在不同的语音层次上影响音节的时长实现；同时，它们又彼此相关，较低层次受较高层次的制约，从高层次的句子一直到低层次的音段，以自上而下的方式发挥作用。

由于实验材料有限，本文提出的一些看法只是对于相对关系方面某些有规律的现象的初步认识，进一步的认识还有待于更加深入的考察。

ABSTRACT

The major studies on coarticulation in Standard Chinese(SC) have been devoted to the spatial dimension of articulatory movements by observing and measuring to acoustic records of relavent segments or syllables. It is still lacking of consideration on temporal dimention, namely, the timing of speech, though some previous work involved in segment duration have been reported.

This paper refers to a study on temporal dimention in natural SC speech, the particular attention is paid to derive a dynamic model of syllable duration in SC by examining their durational variations due to context-sensitivity, i. e. to examine what the main control factors are and how the syllable's duration is affected by these factors.

Our experimental result reported here is mainly obtained from the test materails uttered by one female and two male native speakers. It leads to the considerations that the temporal realization of a syllable in SC is governed by multiple intrinsic and extrinsic factors. The intrinsic ones relate to syllable itself and, besides those phonetically motivated factors, they are mainly coming from stress contrast and tonal difference; the extrinsic aspect is more related to linguistic restriction, besides those universally acted in many other languages, such as the differences in speech rate, syllable position and units' length, the special ones In SC are related to morphological constraint of speech units, difference in syntactic contexts, and the difference in semantic requirements of sentences. These factors are hierarchically worked on different speech levels in a top-down way, the role played in the lower level is usually governed by those on the higher levels. Thus, a temporal model for certain sort of syllable could be predicted according to the intrinsic and extrinsic factors that the syllable is affected by. Consequently, The model offered here is a relational invariant pattern, i. e. a hierarchically dynamic range, rather than a static invariance. So it may be of quite benefit to speech synthesis and recognition, as well as to the basic study on speech timing in SC.

1. INTRODUCTION

Generally, coarticulation involves both of spatial and temporal dimentions of speech, and primarily refers to timing of articulation (Fowler, 1980; Keating, 1988). In nowdays, research on speech timing is greatly concerned with following two aspects: The one is the temporal coordination on the articulatory overlapping and influence occurred between adjacent segments or syllables, it needs intensive approach to the spatial variability along with the passage of time; the other is the adjustment related to durational events themselves, it needs observing and measuring to durational variations due to contextual sensitivity. As a matter of fact, of course, it is impossible to observe the dynamic variations in spatial dimention without reference to the passage of time, and vice versa. Consequently, this two aspects have their own focus respectively, but are closely related to each other. The present study is more related to the later aspect.

Individually, segment's or syllable's duration in natural speech has a wide range of elasticity. It is typically

true in Standard Chinese(SC), from some previous works (Wu, 1986; Cao, 1989a), we found that the maximal value for the syllable uttered moderately in isolation can reach 600 ms and over, and the minimal ones found in connected speech even less than 100ms. Nevertheless, the speech tempo is usually kept in a regular range. Some evidences (Kohno, et al., 1990)in either production or perception from different languages indicate that there may be a universal basis of timing control commonly existed throughout the languages in the world. In other words, there may exist some relational invariance in speech timing, by which the speech tempo is controlled. To search such an invariance, it is necessary to do some intensive work in durational measurements to different languages.

In works on temporal aspect of speech, some control factors affected to the duration of speech sounds have been found commonly from different languages, such as intrinsic phonetic property of segments; speech rate and units' length; position effects in utterance groups and the compensation effects between adjacent units, and so on (Campbell, 1990;Edwards, 1988; Thorsen, 1987; Major, 1981).

Recently, the observations in respect to Standard Chinese also confirmed above findings and made a few improvements. For example,the phonetic property, namely, the articulatory manner of consonants does strongly affect to their durational value, several statistical analyses on the isolated consonants' duration in SC do show the same order as follows (Qi, & Zhang, 1982; Wu,1986):

fricatives > aspirated affricates & stops >
unaspirated affrictes & stops > laterals & nasals

accounting for the obvious loss of closure duration in the measuring to isolated tokens, however, both Feng (1985) and Cao(1989a) had a further measurement, and results in a revising to above sequence by exchange the order of the first and second iterms, that is: aspirated affricates and stops should come first, and then the fricatives.

Considering of the important status of syllable in SC, the present investigation is concentrated on the study of syllable duration, the particular attention is paid to search the relationally invariant patterns of syllables in different contexts by observing the dynamic range of durational variation and examining what the main control factors are and how the syllable's duration is affected by these factors. The preliminary hypothesis is that the durational model of a syllable may be predicted from a relatively invariant framework, which will be derived according to the contextual variations determined by multiple factors.

2. LANGUAGE BACKGROUND AND TEST MATERIALS

This investigation involves two sets of test materials, which were uttered by one female and two male native speakers. The first set consists of 14 pairs of trisyllabic words, which were read in isolation; and the second set contains 6 pairs of sentences, in which the target syllables, those have been already tested in the first set, are appeared as monosyllabic words or as the part of polysyllabic words or phrases. The measurements were made according to the spectrograms of these materials. Considering of the articulatory overlapping occurred between adjacent syllables due to coarticulation in connected speech, the measurement of the overlapped period is counted in both sides of adjacent syllables.

The test materials described above were designed according to following background. First, syllable, as one of the basic units in Chinese speech, is usually running in some relatively fixed groups, like in polysyllabic words or phrases. Therefore, the durational realizations of individual syllable in larger structures may be not only determined by their position in the sequence, but also by the morphological constraint of the structure. In SC, the minimal ones among the fixed groups are the bisyllabic words. This kind of words inherently includes two types of stress, namely, normal type and neutral type. In the former type, both syllables with normal stress; and in the later ones, the first syllable has normal stress and the second syllable is neutralized with a weaked stress. Moreover, the bisyllabic words in SC constructionally have relative stability and strong power in constituting new words, so all of the other polysyllabic structures in SC are essentially the different combinations biult from monosyllabic and bisyllabic words. Therefore, the durational distribution for polysyllabic structures should be closely related to such a constructional feature. To simplify our analysis, here we just take trisyllabic words as the test example.

Secondly, syllable as a basic conveyer for the syntactic and sementic informations, its durational variability may be also affected by syntactic and sementic constraints of sentences. Thus, the test sentences were designed.

In addition to these, for the convenience of discussion, the durational measurements also extends to the acoustic records of monosyllabic and bisyllabic words which were made in our previous investigations (Wu, 1986; Cao, 1989b).

3. EXPERIMENTAL RESULTS AND ANALYSES

3. 1. Average duration

Generally, for the convenience of comparison, people are used to ask for an average value for certain kind of measuremants, however, as a matter of fact, it is hard to assign such a scale for syllable's duration, since in real speech, it varies depending on multiple conditions and usually has a very wide dynamic range. Table I offers some examples of the measurements obtained in this study. According to the figures shown in this Table, we can get an outline of the variable extent that the individual syllable may reach, and realize how wide the dynamic range is, though it is obtained from limited materials.

Table 1. Average duration (ms) of different variations for the syllables /dan/, /jie/, and /shi/ in normal stress utterance

表 1. 正常重音音节‘胆’‘结’‘石’在不同条件下的平均时长.

position		in moderate speech			in faster speech		
		/dan/	/jie/	/shi/	/dan/	/jie/	/shi/
final		332.5	357.5	390	245	290	260
initial		267.5	268.8		222.5	263.3	
inter-	*(a)		193	175		185	135
vocalic	*(b)	187		148	161.3		122.5

*(a) 2nd syllables in trisyllabic words;
*(b) 3rd syllables in quadrisyllabic words.

On the other hand, however, some tendency is observed here, that the number of syllables uttered per second is regularly between 5 — 6 in moderate speech and 6 — 7 in faster speech. If count in the period of intersyllable's overlapping, which is around 30—50 ms according to our observation from this test, then the average duration for a syllable could be estimated at a range of 200—250 ms. By comparison, it is not only far shorter than those tokens in isolated monosyllabic words, in which it is usually over 500ms in moderate tempo, but also shorter than most of the tokens appeared at initial or final positions in connected speech. In most cases, we don't think this average duration has any practical value. Actually, we don't attempt to introduce this figure for any practical use. What we try to indicate here is that, the average number of 5—7 syllables uttered per second seems not to be an accidental event, but rather a reflection of the existence on relational invariance in speech timing. Several relavent studies, either from speech production or from perception, have made great efforts in this aspect. For example, in work on the nature of timing control in languge, Miller G. (1956) suggested that the memory span of holistically processed syllable sequence is about 7 plus /minus 2. Kohno M. and Tsu Shima (1989) confirmed Miller's suggestion according to the evidence that, in child babbling or one word utterance, the syllables in succession never continue over 7. Based on these studies, Kohno M. and Tomoko Tanioka (1990) claimed that the syllable sequence perceived holistically may constitute a 'perceptual sense unit' in process of listening comprehension, and the unit should be a grammatical unit which consists of 7 or less syllables and convey one unit of meaning. Thus, they continued to infer that this unit may offer a universal basis of timing control commonly exists in all the languages. Consequently, our data

obtained from acoustic events tested here may be served as an evidence from Chinese to support such an inference, it is because of that, speech perception is closely related to speech production, so they must share the same basis on timing control. Based on these findings and observation from different languages, we would infer that the marginal seven or less of syllable number may reflect one of the relational invariances in speech tempo existed throughout different languages. Thus, it may be assigned as a relative scale for the comparison of syllable duration.

The figures given in Table I also briefly reflect some effects from the other factors. However, the relationship is more complex than what is appeared here, it will be discussed in later part of this paper.

3. 2. Intrinsic difference

As the situation has been found in many other languages, the intrinsic difference of syllable duration is existed in SC, some kind of syllables do inherently longer than other ones, it is dependent on some intrinsic factors related to syllable itself. However, most of this control factors are language specific, only those physical constraints, such as the differences in articulatory manner of consonants and the hight of vowels, are universally effecient in different languages.

The intrinsic difference found in SC is partly coming from those physical factors. For example, if a syllable with an aspirated initial and a high vowel final or a final with nasal coda, it is likely longer than the others ceteris paribus. On the other hand, however, this difference in SC is also conditioned by some phonological features. Firstly, the tonal contrast is an important aspect affected to syllable duration, usually, a syllable with the third tone, i. e. the mid-falling-rising tone is longer than those with other tones, and the fourth tone, i. e. the high-falling tone is considerably shorter than the others.

Secondly, the inherent difference of stress type is another factor related to syllable duration. Essentially, there is two types of stress related to syllable itself in SC, namely, normal type(NM) and neutral type(NT). In the NM case, the syllable usually has normal sound quality and tone patterns, and a longer duration is expected as well; while in the case of NT, its sound quality is reduced, the tone pattern is neutralized, and a shorter duration is accompanied. Some studies reported that the durational ratio of NT to NM is about 1:2 to 3:5 (Lin, M. & Yan, J. . 1980; Lin, T. 1983; Cao, 1986) Actually, these figures are including some systematic difference, a further test has found that when the neutral type's bisyllabic words are uttered in isolation, the average ratio of NT syllable to its preceding NM ones is around 63:100; while it is around 50:100 when the word occured in connected speech (Cao, 1989b).

3. 3. Contextual variations

3. 3. 1. Variations in isolated bisyllabic and trisyllabic words

As what have been mentioned before, syllable is usually running in groups as a part of larger speech units, so

the durational realization of individual syllable is context-dependent. If we regard the duration of isolated syllable as its underlying form, then that of contextual variation is its surface form. Figure 1. gives the specific surface forms of syllables occurred in different type of isolated polysyllabic words. It is obvious that, these durational patterns are even more complex, we will discussed them in the related part under 3. 3. 3.

Here, let's look at the picture first, there is a dark bar on the far left in the figure, it represents a relative scale for the comparison among the surface tokens, its value is taken from the average duration of the word-initial syllables, and it is very closed to the 'Average Duration' described in 3. 1. The patterns in A and B represent the basic duration model summarized from isolated bisyllabic words obtained from previous investigation(Cao, 1989a; 1989b). It can be seen that in pattern A, the first syllable is equal to or slightly longer than the scale, and the last syllable is clearly longer than the scale; and B is a pattern for neutral type bisyllabic word, in which the first syllable is obviousely elongated and much more longer than the scale, while the last syllable, the NT type syllable, is considerably shortened and far shorter than the scale.

All of the rest bars in this figure are the surface variations occurred in different category of trisyllabic words tested in this study. Here the bars with oblique lines represent the NM type tokens, and those with cross lines represent the NT type tokens. In each case, the shadowed part of the bar shows a scope that a syllable's duration is exactly or likely reached, and the part within dotted line illustrates the extent that is possiblly reached in certain position.

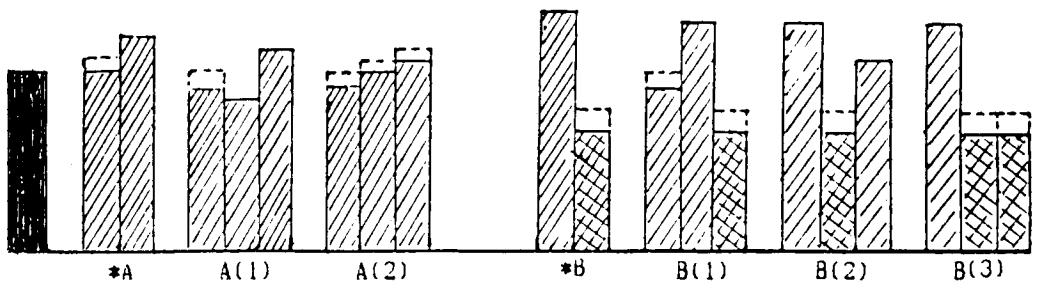

Figure 1. The durational patterns of syllables summarized from vorious isolated polysyllabic words:
*A: NM type bisyll. word; A(1): 2+1 NM type trisyll. word;
A(2): 1+2 NM type trisyll. word.
*B: NT type bisyll. word; B(1): 1+2 NT type trisyll. word;
B(2)&(3): 2+1 NT type trisyll. word.

图 1. 音节在单念的多音节词里的时长模式.
*A: 正常重音型双音节词和A(1):双单式及A(2)单双式三音节词的时长模式;
*B: 轻声型双音节词和B(1):单双式以及B(2)和B(3): 双单式带轻声音节的三音节词的时长模式.

3. 3. 2. Variations in connected speech

So far, the syllable's duration described above is mostly involved in isolated words. In real speech, all of them are organized in connected string. Consequently, their temporal realization must vary in more complex and wider background, the situations have been partially shown in Table I. The average value listed in this table were measured from the variations of syllables /dǎn/[tan], /jié/[ɕie] and /shí/[ʂɿ], which are occurred in different contexts of the test sentences. If make a simple comparison on duration among the three syllables, the intrinsic difference phonetically motivated is still dimly visible, but it has been considerably modified by multiple extrinsic influences.

3. 3. 3. Extrinsic control factors and relavent effects

From figure 1. and Table I, some effects could be clearly observed.

(a) Position effect:

Segments' or syllables' duration is usually affected by their position in the utterance, such as word- /phrase-initial shortening, word- /phrase-final lengthening, and the intervocalic tokens are usually shorter than the average. It is a rule universally acted in many languages. This effect is also found in SC, the figures in Table I have shown a very sharp outline on this phenomenon. Similarly, this tendency also can be seen from Figure 1, for example, the syllables at word-initial in the patterns of A, A(1) and A(2) are all shorter than those at word-final positions. However, the position effect is limited, the detail account is given below.

(b) Morphological constraint:

As the bisyllabic words constructionally have their stability, their durational patterns are often kept even when they occur in larger units. In other words, this bisyllabic structure is always appeared in global. Consequently, as what shown in Figure 1. the durational pattern of A(1), i. e. of 2 + 1 type (compounded by a bisyllabic and a monosyllabic morphemes) trisyllabic word, differ from that of A(2), i. e. of 1 + 2 type's. Obviously, it is due to different placement of the bisyllabic morpheme in each case: in the case of A(1), the bisyllabic structure is globally appeared as the first part of that word, so it is shortened in general, and the following monosyllabic component, as the last part of the word, is lengthened clearly; while in A(2), the bisyllabic structure is appeared as the last part of the word, the constitutional relation in this case is just in reverse comparing with that of A(1). Thus, the different durational pattern is reasonablly resulted. Moreover, this situation is appeared even more prominent in the cases of B(1), B(2) and B(3), where the long-short duration model of the neutral type's bisyllabic structure is kept in any position within different trisyllabic words. In these cases, the second syllable in B(1) and the first syllables in B(2) and B(3) have not been shortened as usual as expected, but obviously lengthened instead, though all of them occur at the intervocalic position or word-initial position. On the contrary, the last syllables in B(1) and B(3) seems have not been lengthened as well, though all of them are occurred at the word-final. In sum, all these phenomena illustrate that the power of position effect is quite limited, its role must be controlled by morphological restriction. In the other words,

position influence can not transgress the effect of morphological construction.

(c) Syntactic context:

In the investigation respect to bisyllabic words in SC, we have observed some influence come from syntactic aspect, for example, the duration of the last syllable in these words are directly affected by the following syntactic context (Cao,1986b). It may be helpful to present an example of the measurements here: when the test word is followed by an independent syntactic element in a sentence, the average duration of the last syllables in these words is around 292 ms; while it is compressed to 254 ms when it is followed by an auxiliary word. Since in the sentence of Chinese, the auxiliary words are usually not independent syntactic element and have to be attached to their previous words, so that the last syllable of previous words become a morpheme located in intervocalic position, instead of the word-final position. Consequently, a shortening effect is resulted due to the position effect. Of course, here is just taking an example to indicates such kind of influence, further study is needed.

(d) Units' length and speech rate :

The other factors affected to syllable's duration are the differences in units' length and speech rate. If we look at back to the situation shown in Table I, and make a simple comparison between first row and the last two rows respectively or that between last two rows, it will be found that, the syllable duration is decreasing with the increasing of syllable number in certain unit. At the same time, the relationship of inverse proportion between syllable duration and speech rate is also found from the difference appeared between 'moderate' and 'faster' columns in this Table.

Notice that, however, there is a phenomenon can not be neglected, that the effect described above is not a simple and linear relation, because regular speech tempo is not only acheived through the elongation or compression of every syllable, but also adjusted by the presence /absence and longer /shorter of possible pauses occurred at phrase boundary. Figure 2 provides a pair of spectrograms for the sentences uttered in different speed. From the picture, about 600 ms difference between this two sentences could be measured. It is obvious that, however, the most prominent difference in this case is about450 ms long pause found after the first phrase in the 'moderate' sentence, but not in the 'faster' one, and the corresponding elongation or compression of individual syllable in either cases does not perform linearly, but only show a tendency in gross.

(e) Variability controlled by semantic requirement:

Speech sound is the surface output represented the underlying semantic input, so the surface realization of syllable duration must be also closely related to semantic requirements of speech. It has been confirmed by the evidence observed in this study. The experimental results obtained from designed test sentences indicate that, this role is usually carried out through different stress accent. Figure 3 gives one of the examples, in which four different statements are made from the same syllable sequence, but uttered in different accent stress according to different underlying * semantic contents respectively, and result in different distribution of syllable duration.

As what has been illustrated in Figure 3, different semantic contents of the sentences result in different placement of the accent stress for every sentence, this effect is usually appeared first in the phrase level, and then the word and syllable levels. For example, statement (1) is a general description, which answers the question

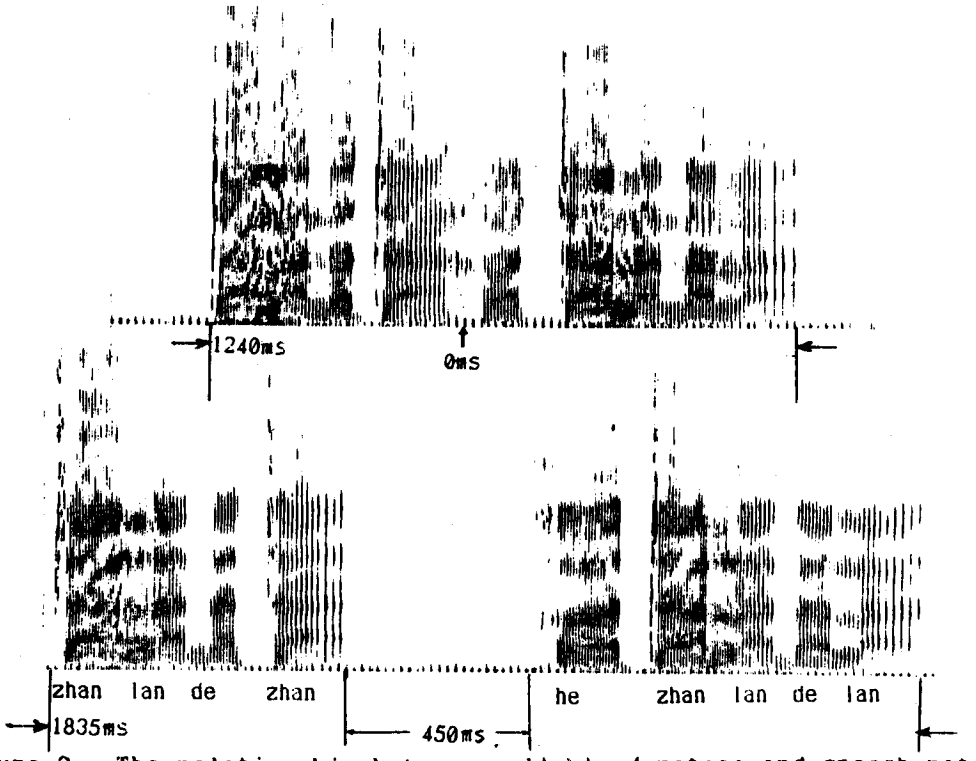

Figure 2. The relationship between syllable duration and speech rate.
图 2. 音节时长与语速的关系示例.

of 'how about this pair of shoes? ', the possible answer could be (not) 'durable', 'siutable' or 'nice' and so on, and the sentence stress is usually located on the predicate part, here is on the phrase of /bu jie shi/, in which, /jie shi/ is a key word and constructionally is a neutral type bisyllabic word. According to word stress rule in SC, here the stress must be placed on the first syllable. Therefore, the accent stress of this sentence is logically set on the syllable /jie/; The situation in statement (2) seems to be similar to (1) in phrase level, but different from (1) both in word and syllable levels, since the underlying semantic content of statement(2) is try to answer the question of 'do this pair of shoes durable?', so the answer emphasizes the negative word /bu/. On the other hand, statement (3) and (4) is similar to each other in phrase level, but different from either in the word and syllable levels, and both of them differ from (1) and (2). In the cases of (3) and (4), what being emphasized is the subject part of the sentence, so the accent stress of the sentence is shifted respectively to the syllable /xie/ or /zhe/ according to their different semantic content. Consequently, the relavent syllables are obviousely elongated. Just look at the percentage summarized bellow, which is the durational ratio of negative word /bu/ to the whole sentence in different statements, we will realize how strong the semantic effect is.

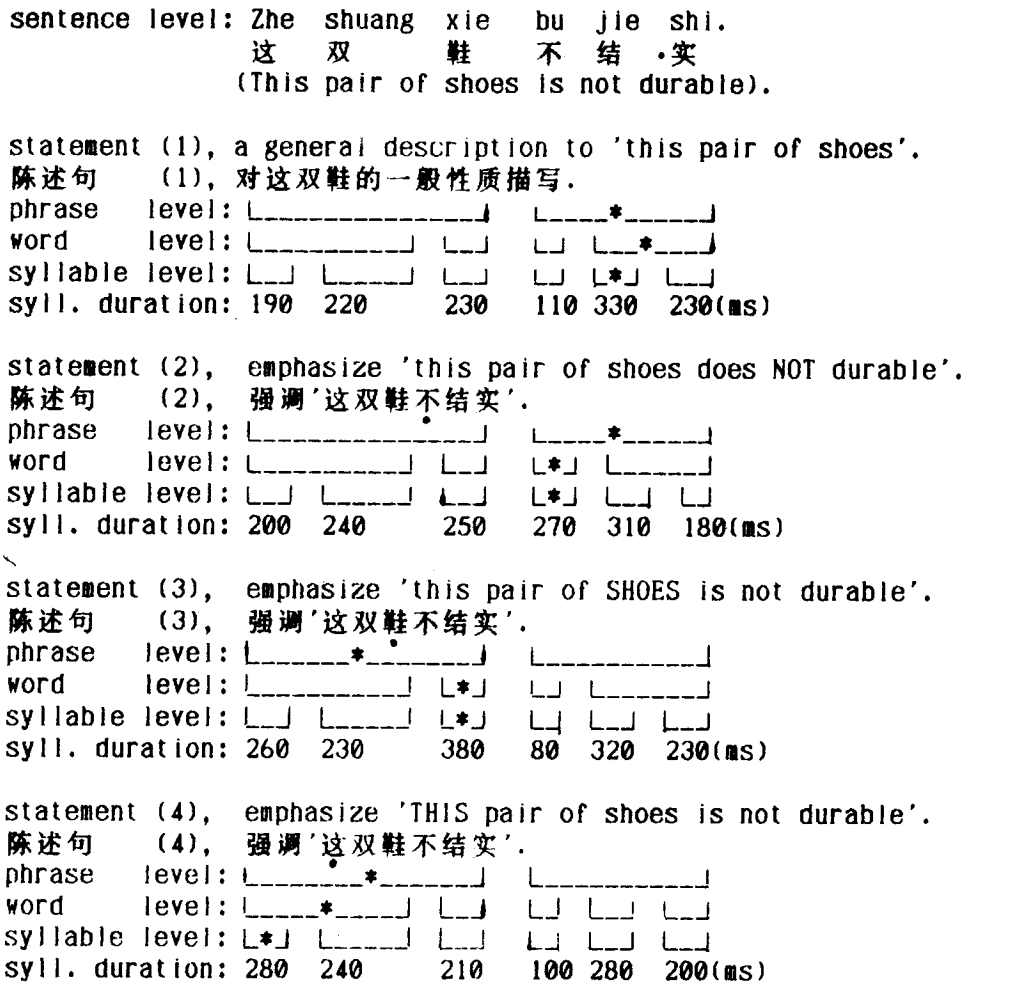

Figure 3. The schematic digram for the relationship between syllable duration and semantic requirement of the sentences.

图 3. 音节时长和句子语义要求的关系示意图

durational ratio	statements
9.3%	(1)
21.4%	(2)
5.4%	(3)
7.8%	(4)

4. CONCLUSSION

Based on the limited observation taken in this study, the durational patterns of syllables in Standard Chinese could be summarized as follows:

(a) Syllables as monosyllabic words moderately uttered in isolation is usually over 500 ms. Intrinsic difference is existed among them, which is conditioned by intrinsic factors related to syllable itself. However, all of these isolated tokens will be modified by different extrinsic influences. First, it will be much more shortened when they ocuur even as an independent elment in connected speech, the average ratio between this two kind of tokens is around 100:70; Second, it will vary depending on multiple factors and run in some relatively fixed patterns. Therefore, the duration model offerred here, as it has been shown in the table and figures, is relationally invariant patterns, instead of static invariance.

(b) Syllable duration in real SC speech has a wide dynamic range: the higher threshold is about 600 ms, the lower one is around 100 ms, and the specific realization is appeared so dazzling. However, as what has been found in many other languages, the speech tempo is still kept within a regular range, it is achieved through the mechanisms of intrinsic compensation and adjustment. In different conditions, the specific elongation or compression of individual syllable undergone is not linear, but disproportional.

(c) The control factors affected to the duration realization of a syllable are mainly coming from two aspects, namely, intrinsic and extrinsic factors. The intrinsic ones are phonetically and phonologically motivated, which is related to syllable itself including articulatory manner of initial consonants, tonal contrast, and the difference in stress type, and so on. The extrinsic factors are more related to linguistic aspect, it contains speech rate and units' length, syllable position and syntactic context, morphological constraint and semantic reqirement. All these factors are hierarchically worked on different speech levels in a top-down way, the role played in the lower level is usually governed by those on higher levels. Thus, the durational pattern for certain sort of syllable could be predicted from the relational framework described above according to the factors that the syllable being controlled by.

REFERENCES

Campbell, W. N. (1990). Evidence for a syllable based model of speech timing. The Proceedings of ICSLP 90, 1. 3. 1—4, Kobe, Japan.

Cao, Jianfen, (1986). Analysis of neutral tone syllables in Standard Chinese. Applied Acoustics Vol. 5, No. 4: 1—6.

Cao, Jianfen, (1989a). Analysis on temporal distribution of bisyllabic words in Standard Chinese. Proceedings of the 4th National Conference on Speech, Image, Communication and Signal Processing, 84—87, Beijing, China.

Cao, Jianfen, (1989b). Temporal distribution of bisyllabic words in Standard Chinese: An evidence for relational invariance and the variability from natural speech. The RPR—IL (CASS)/1989, 38—56.

Edwards, J. & Beckman, M. E. (1988). Articulatory timing and the prosodic interpretation of syllable duration. Phonetica 1988, 45, No. 2—4: 156—174.

Feng , L. (1985). Duration of initials, finals and tones in Beijing dialect. Working Papers in Experimental Phonetics, 1985, pp. 131—195.

Fowler, C. (1980). Coarticulation and the theories of extrinsic timing. Journal of Phonetics 8:113—133.

Keating, P. A. (1988). Coarticulation and timing. UCLA Working Papers in Phonetics 69: 1—2.

Kohno, M. & Tsu Shima, (1989). Rhythmic phenomena in a child's babling and one-word sentence. The Bulettin No. 191: 6—13, The Phonetic Society of Japan.

Kohno, M. & Tomoko Tanioka, (1990). The nature of timing control in language. The Proceedings of ICSLP 90, Kobe, Japan.

Lin, M. & Yan, J. (1980). Acoustic characteristics of neutral tone in Beijing Mandarin. Fangyan, No. 3, 166—178.

Lin,T. (1983). A preliminary exploration to the nature of weak-stressed syllable in Pikingness. A collaction of articles in linguistics, Vol. 10, The Department of Linguistics at Beijing University.

Major, R. C. (1981). Evidence of stress timing in Brazilian Portuguese. Journal of Phonetics 9: 343—351.

Miller, G. (1956). The marginal seven, plus or minus two? Psychological Review 63—2, 81—97.

Qi, S. & Zhang, J. (1982). An analysis of consonant duration in Standard Chinese. Acta Acoustica Vol. 7: 1.

Thorsen, N. G. (1987). Suprasegmental transcription. Probleme Der Phonetischen Transkription, edited by Herausgegeben Von, Antonio Almeida und Angelika Braun, Stuttgart, 1987, pp. 84—87.

Wu, Z. (1986). The Spectrographic Album of Monosyllables of Standard Chinese. Chinese Social Sciences Press, Beijing, 1986.