

TEMPORAL STRUCTURE IN BISYLLABIC WORD FRAME:  
AN EVIDENCE FOR RELATIONAL INVARIANCE AND VARIABILITY FROM STANDARD CHINESE

Cao, Jianfen

Institute of Linguistics, Chinese Academy of Social Sciences  
5 Jianguomennei Road, Beijing 100732, CHINA

ABSTRACT

The effort of this paper is concentrated on searching for relational invariance in temporal distribution of the bisyllabic structure in Standard Chinese by observing the durational variability of bisyllabic words in different conditions with regard to context, position, stress type and so on in real speech. Experimental results indicate that both of durational variability of bisyllabic words and temporal distribution within the word in Chinese are determined by multiple influences from either higher or lower layers, however, the influence from morphophonemic contrast of word-stress is the most powerful one, and the other influences must be governed by such morphophonemic constraint. In this sense, temporal structure in Chinese is word-stress dominated, and as relational invariance, such word-stress based temporal patterns are relatively stable and kept as a fixed group in the continuous speech through intrinsic adjustment and compensation. Considering of the special status of bisyllabic words, this kind of patterns may be served as a miniature on timing of real speech in Standard Chinese.

1. INTRODUCTION

It has been a matter of great urgency to resolve the discrepancy between static linguistic units and dynamic speech processes[1]. Since in natural speech the realistic instances of sounds are so complex and there are too many variants to recognize the original patterns as usually appeared in the classical literatures. Therefore, experts with deep insight have claimed for the urgent need of 5th generation's linguists [2] to pay more attention to the approach on relational invariance and variability. Many contributions have been made, though some opinions still being controversial[3]. Practically, it is worth to do to search for some rules and patterns from realistic speech sounds, these patterns should be less in contextual sensitivity or is certain factor-dependent, so that it can meet the demands of highly developed science and technology.

In comparison with the study on spatial dimension in Chinese, the approach in time dimension seems even further behind the fashion, though a number of investigations [4] [5][6][7][8] have contributed recently. The present study would provide a preliminary investigation to the temporal distribution of bisyllabic structure in Standard Chinese by measuring the durations of bisyllabic words and each subpart in question within the word. The effort is first concentrated on description to the effects of certain factors upon the durational variability at word level, and secondary on a brief observation to the temporal distribution of intrasyllabic and intersyllabic cases within the word, both of them serve for the same aim that to find out whether there any regular time patterns exist in natural Chinese speech.

The hypothesis is that the temporal distribution of bisyllabic structure should follow some regular patterns, which may be dependent on multiple factors, but dominant one is the contrast of morphophonemic stress. Consequently, such kind of word-stress based temporal structure should be relatively stable.

II. LANGUAGE BACKGROUND AND TEST MATERIALS

The language investigated in this study is Mandarin as the Standard Chinese (SC).

Bisyllabic word is the most important structure with the highest occurrence and strongest power in building new words in this language. Moreover, it is the minimum unit as connected speech. Therefore, as the investigated object, they can cover all of the temporal patterns for either intersyllabic and / or intrasyllabic cases within a word, so it will be of inevitably quite benefit to the study on temporal structure of connected speech to firstly clarify the situation in bisyllabic words.

Morphophonemically, bisyllabic words in SC inherently include two types of stress, namely, normal (NM) type and neutral (NT) type. The former type consists of syllables both with normal stress, i.e, with normal duration, sound quality and tone patterns; while in the latter type's, the first syllable has normal stress, but the second one is neutralized with weak stress, i.e, has shortened duration, reduced sound quality and neutralized tone patterns.

According to the language background described above, the test materials involved in this study contain two groups of bisyllabic words. The first group consists of eight NM words and twelve NT words. For looking at the possible effects of differences in syllable position or stress type upon temporal distribution, all the words chosen in this group are reduplicative ones, so that the other conditions could be kept in constant. The another group of words contains nine pairs of contrastive words. Each pair of words in this group share with the same phonological structure but differentiate in stress type, so that the words in this group is minimally contrasted in their stress type.

To examine possible contextual influence of the higher level upon test words in connected speech, two different styles of carrier sentences are employed for the words of group 2. They are listed as follows:

Chinese transcription	gloss
a) wo shuo * * zhe ge ci.	I say the word * *.
b) Ni shi shuo * * ba?	Did you say * *?

All the test words were read in isolation at normal speed by one male and one female speakers; the second group of words were first read in isolation and then in different carrier sentences by the same speakers.

III. RESULTS AND DISCUSSION

The results can be summarized in the Tables(see p.2).

Individually, the duration of certain speech unit in SC has very wide range of elasticity, which is determined by various influences, which is from either higher and/or lower layers. The relevant approach[7] has found that the temporal realization of a syllable in SC is governed by multiple intrinsic and extrinsic factors. The intrinsic ones is related to syllable itself, that is phonetically or phonologically motivated; while the extrinsic aspect is more related to linguistic restriction, specifically, it is related to the constructional restriction of speech unit, the differences in syntactic contexts and underlying

10.21437/ICSLP.1992-355

semantic requirements of a sentence. These factors are hierarchically worked on different speech levels. Therefore, it must be also effective to all the speech units other than syllables.

Based on the experimental data listed above, some basic situations can be observed from different levels.

Table I Average duration(ms) of bisyllabic words in isolated(IS) and connected(CN) speech

	NM words			NT words		
	token	mean	sd.	token	mean	sd.
IS	16	609.14	67.52	16	538.3	46.10
	*A. 18	586.64	77.94	18	498.75	60.75
CN	*B. 18	497.93	77.19	18	453.15	63.93
	plotted. 36	542.28	62.27	36	475.95	32.24

\*A. Followed by an independent structure  
\*B. Followed by a dependent function word

Table II Average duration(ms) and the ratio(%) occupied by first and second syllable in bisyllabic words

	in NM words				in NT words				
	syllable	token	mean	sd	ratio	token	mean	sd	ratio
IS	1st	16	278.92	63.49	46	16	329.61	45.36	61
	2nd	16	330.22	64.79	54	16	208.69	29.4	39
CN	1st	18	296.33	40.81	55	18	322.13	47.99	68
	2nd	18	245.95	49.15	45	18	153.82	31.15	32

Table III Average duration and its ratio of initial consonants in first syllable of the words

initial type	word stress	no. of token	initial mean(ms)	syllable mean(ms)	initial ratio(%)
aspirated stops & affricates	NM	6	121.72	282.65	43
	NT	10	138.53	333.96	41.5
fricatives	NM	14	125.00	347.48	36
	NT	16	136.19	340.49	40
unaspirated stops & affricates	NM	20	78.36	301.31	26
	NT	22	92.66	309.70	30
nasals & laterals	NM	12	69.96	281.25	25
	NT	12	83.96	307.84	27
plotted	NM	52	98.76	303.17	32.5
	NT	60	112.84	323.00	34.9

### 3.1 Durational variability in word level

According to the data shown in Table I, the duration of bisyllabic words seems mainly vary depending on their position in speech, the difference of syntactic context in sentence and stress contrast of the words themselves.

#### 3.1.1 Influence in the contrast of word-stress

Generally, the bisyllabic words in SC are appeared in global as relatively fixed unit in terms of their stress type. According to the relevant approach(7), the temporal realization of a syllable is sensitive to its position in polysyllabic structures, however, this position effect is limited, and as a fixed unit, the bisyllabic frame are kept in any position. Consequently, the manifestation of word duration should be governed by such a constitutional stability, too. The result listed in Table I has revealed that there does exist a systematic difference between the NM and NT type words, no matter it is in the case of isolation or connected speech, the NT type word is always shorter than the NM type's. According to the data obtained here, the NT word is about 12 percent shorter than NM word.

#### 3.1.2 Effect of contextual sensitivity

Usually, duration of the word in isolation is longer than that in connected speech, this systematic difference

can be observed by comparing the measured data shown in Table I. Specifically, however, there is more complex variants in connected speech due to certain contextual sensitivity. For example, when a word occurs in pre-pause position, the status is similar to that of isolation, where an effect of phrase-final lengthening is occurred, and a longer duration of the word is resulted; Whereas it is located in the post-pause or intervocalic positions, its duration will be relatively shorter and vary depending on the difference of syntactic context.

As what can be seen from the bottom of Table I, the durations of the words in group B is obviously shorter than that in group A. This is due to the difference in syntactic structure between the carrier sentences A and B. Since in Carrier sentence A, the test words is followed by a phrase "zhe ge ci" (this word), which is an independent syntactic element in the sentence. In this case, a longer interval exists at the boundary between test word and the phrase, consequently, an effect of word-final lengthening occurs naturally, and results in a longer duration of the test word. Whereas in sentence B, the test words are followed by an interrogative word "ba", which is an auxiliary word and has to be attached to its previous word as an immediate constituent, and it is impossible to inlay any pause at the boundary between previous word and the function word. Thus, the interval between the test word and the "ba" just equal to an intrasyllabic gap. This situation makes a limitation to the duration of the last syllable in test word, and causes a shortening of the whole word, the illustration given in Fig.1(next page),

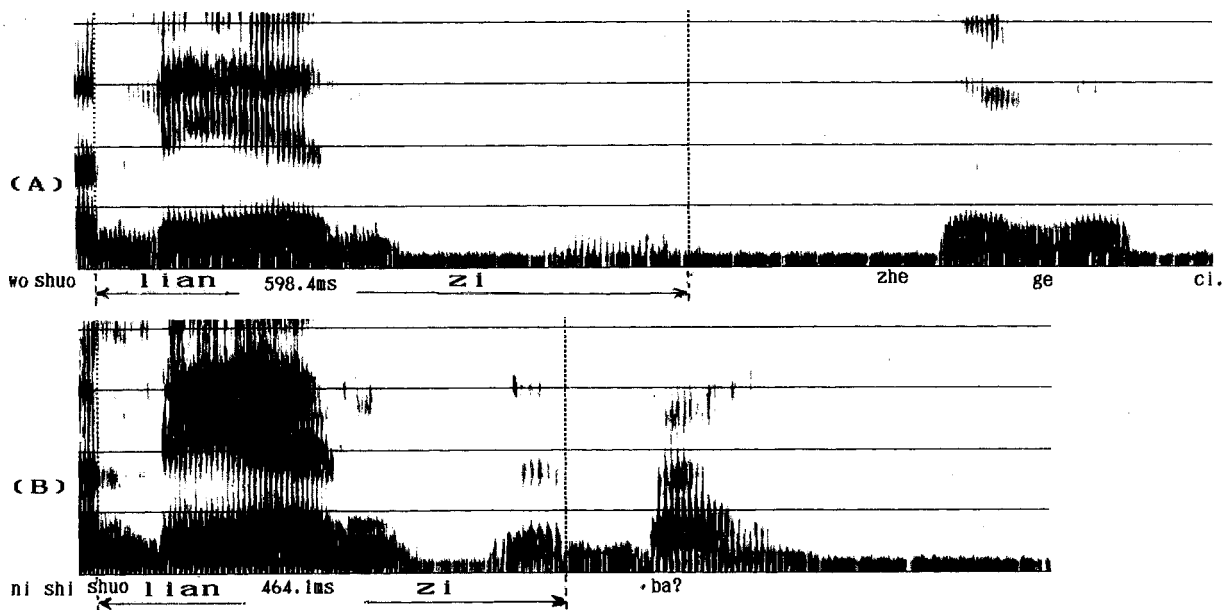


Fig. Example of the difference in durational realization of the test word in carrier sentences (A) and (B).

### 3.2. Temporal structure within the word.

#### 3.2.1 Effect of the difference between IS and CN.

From the data presented in Table II, a tendency can be observed that temporal distribution within a word is obviously conditioned by the contrast of isolation and connected speech. Take the reduplicative NM words as the example, in isolated bisyllabic words, the first syllable is usually shorter than the last syllable. The average duration of first syllables is about 84% of the last ones. However, the situation is just reversed in connected speech, average duration of the last syllables is about 83% of the first one's. This phenomenon indicates that the last syllable of the word is much more shortened due to contextual limitation in connected speech.

The same tendency is found for the NT type words, too. As what has been mentioned before, the last syllables in the NT type words is neutralized and inherently shorter than the first syllable in SC. Actually, such a difference appears even more prominent in connected speech, where the neutralized syllables are further shortened, as the mean value shown in the right column of Table II, in isolated words, the duration ratio of last syllable to first one in NT words is about 0.63:1, while it is reduced to about 0.5:1 in connected speech.

#### 3.2.2 Role of word-stress type

There is a tendency can be seen from the durational ratio in Table II, that in either case of isolation and connected speech, the first syllable in NT words is considerably elongated in comparison with that of the NM words. It is because of that, last syllable in NT word is neutralized, and it is inherently much shorter than normal syllable. However, the regular rhythm of natural Chinese speech does not allow the NT words with far cry from the NM words in duration, thus, as a compensation to such a shortening, the effect of elongation is applied to the first syllable, so that to meet the rhythmic demand of natural speech.

### 3.3 Temporal distribution within a syllable

According to the measured data so far, the temporal distribution within syllable seems to follow a relatively

stable ratio. In the normal type syllables, the ratio held by initial is usually under 50 percent when it occurs in NM words; whereas the ratio is always over 50 percent when it stays in the NT words. Of course, this is only a rough range, the specific situation is much more complex. It is dependent mainly on following factors.

#### 3.3.1 Difference in phonetic nature of initials

Obviously, the durational ratio of initial consonant in a syllable is determined by its articulatory manner. According to the data (see Table III) obtained in this investigation, the durational order for different initial consonants can be lined up as follows:

- a) aspirated affricates and stops,
- b) fricatives,
- c) unaspirated affricates and stops,
- d) nasals and laterals.

#### 3.3.2 Stress type of syllable itself

Generally, the initial consonant in the neutralized syllable held greater ratio of duration than that in the normal type one. It is not surprised, since some previous studies (9|110) have found that the final of neutralized syllable in SC is obviously reduced and shortened. Thus, as an effect of automatic compensation and adjustment, the elongation taken place in this case is reasonable.

#### 3.3.3 Contrast of word-stress

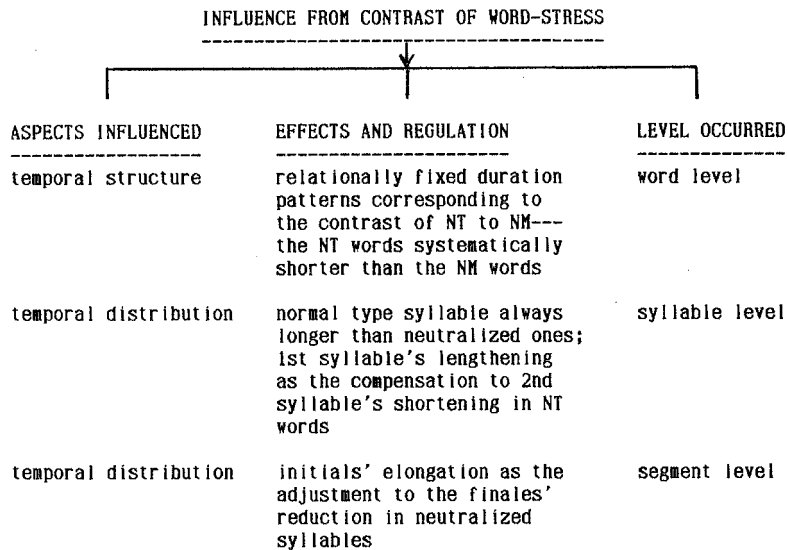
The correlated figures in Table III show that, even in the normal type syllable, the temporal structure is also restricted by the stress contrast of the word where it occurred. In general, either the mean duration or the ratio held by the initials in NT words are systematically longer or higher than that in NM words, except the case for the pair of aspirated initials, it may be caused by the small number of token. However, the relevant analysis of variance indicates that this sort of distributional difference is statistically significant ( $P < 0.01$ ).

So far, the situations described above have provided a strong evidence that the morphophonemic contrast of word-stress in SC is the key factor in the influence upon the temporal distribution of bisyllabic words. As a basic factor, this kind of contrast not only affects to the manifestation of duration in the word level, but also

dominates the temporal distribution of sub-levels within the word, though the specific duration of certain syllable or segment is definitely sensitive to contexts or varied depending on its phonetic or phonological constraint. Thus, the temporal structure is actually word-stress based and the corresponding patterns are kept everywhere through various regulations, among which, the most typical ones

are the compensational lengthening of pre-neutralized syllables and the elongation effect of initial consonants correlating to the reduction of the finals taken place in neutralized syllables.

The relationship described above may be summarized briefly by following chart:



#### IV. CONCLUSION

To sum up, the experimental results obtained from this investigation has supported the hypotheses made at the beginning of this paper. It seems that the temporal structure of bisyllabic words in Standard Chinese depends on multiple influences, such as syntactic difference in sentence, morphophonemic contrast of word, different type of syllable stress, phonetic nature of segment, and so on. Usually, the lower the layer that speech unit stay in, the more the restriction to be endured. However, the most powerful restriction that every unit in different levels has to be endured is the morphophonemic contrast of word-stress. Nevertheless, there do exist some relationally invariant patterns either in durational variability at the word level and the temporal distribution within the word. Basically, the patterns are word-stress based and relatively stable in real speech. Generally, this sort of relationally invariance is executed through the automatic adjustment and compensation taken place in different speech layers.

#### V. REFERENCES

- [1] Kohler, K. J. (1981). Temporal aspects of speech production and perception (preface). *Phonetica* 38:5.
- [2] Fant, G. (1983). Phonetics and speech technology. Proceedings of the 10th International Congress of Phonetic Science, Holland.
- [3] Beckman, M.E. (1988). *Phonetic Theory*, Linguistics: The Cambridge Survey, Vol.1:216-238.
- [4] Qi, S. & Zhang, J. (1982). An analysis of consonant duration in Standard Chinese. *Acta Acoustica* Vol.7:1.
- [5] Ren, H. (1988). Temporal structure of the syllables with nasal ending in Chinese. Annual Report of Phonetic Research, PL-ARPR, 53-79. Phonetic laboratory, Institute of Linguistics of CASS.

- [6] Feng, L. (1985). The durations of the initial, final and tone in connected Pekingese, An Collections of Speech Sounds in Pekingese.
- [7] Cao, J. (1991). The durational patterns of syllables in Standard Chinese. Proceedings of the 12th ICPhS, Vol.2: 310-313. Aix-Provence, France.
- [8] Wang, J. & Wang L. (1991). The durational patterns of syllables in polysyllabic words of Standard Chinese. Report given at the Chinese Phonetics Workshop, Beijing, 1991.
- [9] Lin, M. & Yan, J. (1980). The Characteristics of neutral tone in Standard Chinese. *Fang Yan* 3.
- [10] Cao, J. (1986). An analysis for neutralized syllables in Standard Chinese. *Applied Acoustics*, Vol.5:4.