RHYTHM OF SPOKEN CHINESE
-- LINGUISTIC AND PARALINGUISTIC EVIDENCES--

CAO Jianfen

Institute of Linguistics
Chinese Academy of Social Sciences
Email: Jianfencao@hotmail.com

ABSTRACT

This paper discusses rhythmic characteristics of Standard Chinese. The study is based on relevant acoustic-phonetic investigations, including some paralinguistic analyses. The preliminary result leads to the considerations that (1) there is no evidence to regard Chinese as a syllable-timed language; (2) speech rhythm may be more related to regular occurrence and variation of certain prosodic phenomena at particular positions, in stead of isochronal recurrence of any speech unit.

1. INTRODUCTION

Traditionally, in phonetic theories, rhythm was regarded as a phenomenon related to regular recurrence in time of some given speech unit, and suggested that such "isochrony" is commonly existed in all spoken languages. Thus, the speech rhythm of different languages to be categorized accordingly. For example, it is well known that English is a typical stress-timed language, that means the periodic recurrence of movement is supplied by the stresses-producing process; and French is a typical syllable-timed language, that means the syllables recur at a equal intervals of time (e.g. Pike, 1946). This view has persisted for centuries in phonetic theories, and still is a very pervasive idea up to date.

Respect to this idea, many phonetic studies on rhythm have focused on the search for acoustic evidence for isochrony in order to support the traditional suggestion. Unfortunately, however, these studies have shown that there is no reliable acoustic basis for isochrony either in the inter-stress intervals of stress-timed languages or in the syllable duration of syllable-timed ones, so the suggestion on isochrony can not be confirmed until now.

In some literatures, Chinese to be regarded as so-called "syllabism language", it implies that Chinese is also syllable-timed. Therefore, many people even had simply considered that language processing in Chinese should be much easier than that in other languages, while practical experiences in field have forced them to change their mind, since syllable duration in Chinese has wide range of dynamic variation. The situation is quite complex. Consequently, in spoken language processing, people have to face such unfortunate fact and try to work out a solution. In this background, there are increasing interests in recognizing the essence of rhythm and the timing of Chinese. Thus, to explore the nature of rhythm in Chinese has being an urgent task.

2. RELEVANT INVESTIGATION IN CHINESE

In last ten years, the author of this paper have conducted several investigations in order to examine Chinese rhythm from the nature of speech timing (e.g., Cao, 1991; 1992; 1994; 1995; 1999). In these studies, the main test materials employed were read style speech, including words, sentences and discourses selected from TV news announcement or the prose declaim recorded from broadcasting. Through these investigations, the basic characteristics related to Chinese rhythm has been observed. The main points could be summarized as follows.

2.1. Syllable duration

In natural Chinese speech, syllable duration contains a wide range of variation due to the effects of multiple factors. These factors are mainly coming from two aspects, namely, intrinsic and extrinsic aspects. The intrinsic ones are phonetically and phonologically motivated, it is related to syllable itself including difference of phonological constituent, tonal distinct, lexical stress contrast, and so on. The extrinsic factors are context-dependent, and it is more related to paralinguistic influence that referred to speech emotion, which is deliberately controlled by the speaker, but not motivated by phonetic or phonological constraints.

All the factors mentioned above are hierarchically worked on different speech levels in a top-down way, the role played in lower level is usually governed by those on higher levels. For example, generally the third tone syllable in Standard Chinese is longer than the other tone's syllable, and the fourth tone syllable is the shortest. The syllable with fricative initial and compound final is usually longer than that with nasal initial and simple vowel final. Besides, in various polysyllabic words, the duration manifestation for each individual syllable is obviously conditioned by their lexical stress contrast. However, these intrinsic difference will be further

modified by the extrinsic factors, among them, the most powerful effect is coming from paralinguistic emphasis and the difference of syllable location in phrase or sentence.

2.1.1. The effect of emphasis upon syllable duration

According to the data measured from read sentences, the manifestation of syllable duration is definitely determined by the change of emphasis in utterance, no matter what its intrinsic characteristics is. The amount of duration difference between the syllables of emphasized and non-emphasized is quite significant. Table I shows an example of syllable duration manifested in moderated read sentences, the data were collected from 9 speaker's utterance. In which, the S1, S2, S3 and S4 are consist of the same syllable string of "Zhe shuang xie bu jie shi"(this pair of shoes is not durable), but read out with different location of emphasis. From this table we can learn that, the duration of emphasized syllables is significantly longer than that of non-emphasized partners, and this tendency is identical among different speakers.

Table I Duration of syllables : multi-speakers' mean value(ms) and durational ratio(%) in S1, S2, S3 and S4 where the emphasis is falled on different syllables (the corresponding figure is in bold).

Sentences / Syllables	S1	S1	S3	S4	Mean
Zhe	190 14.5	200 13.8	**280** **21.4**	230 15.3	225 16.3
Shuang	220 16.8	240 16.6	240 18.3	260 17.3	240 17.3
Xie	230 17.6	250 17.2	210 16.0	**380** **25.3**	268 19.0
Bu	110 8.4	**270** **18.6**	100 7.6	80 5.3	140 10.0
Jie	**330** **25.2**	310 21.4	280 21.4	320 21.3	310 22.3
Shi	230 17.6	180 12.4	200 15.3	230 15.3	210 15.2

At the same time, the role of emphasis upon syllable duration is also accompanied by an obvious pitch prominence. It can be seen in Fig. 1(omitted here). In which, both of F0 contour and F0 range are standing out clearly.

2.1.2. The effect of difference of syllable position within rhythmic unit

Generally, there are obvious duration difference between the syllables in different prosodic position, such as those occurred at prosodic phrase-initial vs. phrase-final positions. Usually, the duration of phrase-final syllable is 1.3 times of that of phrase-initial one's in average, while those in sentence end are not lengthened and even shortened in most of the case, the details will be specified later in section 2.4.

2.2. Rhythmic grouping

Generally, during speaking, people neither utter without any break nor syllable by syllable separately, but combining several syllables into a larger rhythmic chunk according to temporary needs of semantic expression. At the same time, in perception, people are also sensitive to such kind of chunks, in stead of individual syllables. Some studies have noticed that this phenomenon is based on human cognitive mechanism (Laver, 1994). It is also evidenced in Chinese.

Syllables in Chinese speech are generally grouped into polysyllabic chunks, among them, the most common and powerful one is disyllabic and trisyllabic chunk. Functionally, this size of chunk forms a foot or prosodic word (hereafter PW), and acts as the basic building-block of temporal organization. One or more such chunks to be grouped into prosodic phrase(PP) and further intonation phrase(IP) respectively. In addition, there are certain coherent features within rhythmic unit and obvious boundary markers between rhythmic units in each prosodic level(Cao, 1999).

2.3. Temporal structure of rhythmic unit

Generally, temporal distribution of rhythmic units in Chinese has relatively fixed patterns. For example, durational distribution within PW is regularly based on lexical stress contrast(as what shown in Fig. 2, omitted here), it is identical to the duration patterns of disyllabic and trisyllabic words (Cao, 1991). On the other hand, however, the duration manifestation for certain rhythmic unit is context-dependent, and especially sensitive to the change of some paralinguistic information, such as the emphasis added by speakers.

Table II shows the duration manifestation of PP and PW in the sentences with different emphasis. The data is obtained from the test materials the same as those mentioned in 2.1.1. Specifically, in S1 and S2, the speaker insists to emphasize the PP of "bu jie shi" and further concentrated on the PW "jie shi" in S1; whereas, in S3 and S4, the main attention is paid to the PP of "zhe shuang xie" and further to the PW "zhe shuang" in S3. Thus, the durational ratio of corresponding PP and PW are standing out clearly. This situation indicates that temporal distribution between rhythmic units is also sensitive to speakers' intention.

Table II Durational ratio(%) of PPs and PWs in the

sentences with different location of emphasis
(the corresponding figure is in bold)

PPs/PWs / Sentences	Zhe shuang xie	Bu jie shi	Zhe shuang	Jie shi
S1	48.9	**51.1**	31.3	**42.7**
S2	47.8	**52.4**	30.3	33.8
S3	**55.7**	44.3	**39.7**	36.6
S4	**58.0**	42.0	32.7	36.7

2.4. Distribution of Pausing and pre-boundary lengthening at rhythmic boundary

The size of pause interval and pre-boundary lengthening at rhythmic boundary are varied depending on the boundary strength. The specifications can be observed from Table III. In which, we can see that, pause duration is in following order: pause between paragraphs> between IPs > between PPs. And the amount of pre-boundary lengthening is ordered as follows: the maximum lengthening occurs at PP-final, but usually no lengthening taken place at the end of sentences and paragraphs, where mostly occurs a shortening.

Table. III Temporal distribution at rhythmic boundary:
A. Lengthening/shortening in terms of durational ratio of the rhymes in pre-boundary syllables ;
B. Duration of silent pauses between PPs and IPs:
a. Sentence end, b. Paragraph end

	A(%)		B(ms)	
Phrase Type / Speech Style	PP-final	IP-final	B/w PPs	b/w IPs
Female Speaker	1.68	0.93	154	a. 538 b. 1112
Male Speaker	1.55	0.86	397	a. 719 b. 2000
News Speech	1.62	0.90	276	a. 629 b. 1020
Declaim Speech	1.33	1.05	59	a. 548 b. 2000

Basically, rhythmic grouping is mostly related to the temporal dimension, however, as a sort of coordination, it is also manifested via pitch register resetting at phrase and sentence boundary. The size of such resetting usually increases with the pause duration.

3. DISCUSSION

Apparently, according to the investigated results summarized above, we not only make no sense to regard Chinese as a syllable-timed language, but also find no base to evidence "isochrony" to be the nature of speech rhythm.

Nevertheless, rhythm is commonly existed in all spoken languages, people have strongly noticed that rhythm is referred to chunking strategy and closely related to the time behavior of speech production and perception. The problem is that what is the matter about that? How to view the relationship between rhythmic grouping and speech timing?

In Some earlier study, rhythmic unit was defined as breath grouping(e.g., Passy 1930), and this term still continue to be used in many literatures up to date. Apparently, this is taking a viewpoint of physiological mechanism, and it may be one source of the idea of so-called "isochrony", because human breathing is taking roughly equal time interval. However, Passy also pointed out that such grouping behavior is also related to sense grouping. Therefore, every breath grouping is roughly corresponding to a simple sense unit. Thus, in speech, the sound grouping is more or less constrained by logical principal, while such turning is varied, larger or smaller, so the interval of each grouping does not have to be isochrony. Moreover, yet he specified that breath grouping can be further divided into grouping of force, in which there must be one syllable is accented. Generally, each of such force grouping is combined by two or three syllables. These syllables are cohered very closed in sense, and there is one among them must be more important in sense, and the force group can also work as a breath group when the speech tempo is relatively slow. Consequently, according to Passy(1930), rhythm grouping not only doesn't necessary to be isochrony, but also seems not possible to be isochrony in most cases.

As what described in section II, in Chinese, there is no evidence to confirm the suggestion of "isochrony", too. However, we do find some interesting phenomena that may contribute to understand rhythmic grouping and speech timing. The one is that the span of syllable number in The PP is varied within a certain range, in other words, it is close to a constant. The other is the timing behavior of rhythmic unit seems to be relationally invariant. These regular phenomena may be result of an isochronic impression. The details will be specified bellow.

3.1. The span of syllable number in prosodic phrase

According to the study respect to Standard Chinese (Cao, 1999), PP in Chinese is the most frequent chunk in speech production and perception. The length of PP is varied all time in speech, however, the variation seems to be limited about 7±2 syllables, especially when these syllables occur in relatively unstressed positions. Besides,

the specific span of such PP chunk is also restricted to the difference of speech rate and speech style. For example, according to the data come from the news speech, the span is mostly around 7 syllables; while that is mostly around 5 in declaim speech.

The length limitation of prosodic phrase observed above may be not surprised, since there is similar phenomena also found in other languages. For example, early in 1978, Boomer reported that "in spontaneous speech, there are discernible 'chunks', sequences of a few syllables, usually from one to seven or eight, that seem to be spoken as a unit". Moreover, the reports from psycholinguistic approach (e.g. Dittmann etc., 1967) suggested that such size of clause "is a plausible candidate for psycholinguistic unit of speech decoding, as well as for speech encoding". In addition, according to some studies in other relevant fields, the memory span of holistically produced syllables sequence is about 7 plus /or minus 2 (Miller, 1956), and the syllables in succession never continue over 7 in child babbling or one word sentence (Kohno, M. and Tsu Shima, 1989). Consequently, I would suggest that, the limitation found here in Chinese should not be an accidental event, but a further evidence for the findings made by Dittmann etc(1967). It may imply a relational invariance related to a common rule on timing control in speech production and perception, hence cause an perceptual impression of appeared "isochrony".

3.2. Temporal adjustment in phrase level

Table IV shows the situation on durational distribution of phrase-initial and phrase-final syllables. From the data listed in this table, a regular adjustment in speech tempo can be seen clearly. First, the duration of phrase-initial syllable is systematically different from that of phrase-final syllable; second, this difference clearly conditioned by the type of phrase. Specifically, in the phrase that to be continued in sense, the duration of the first syllable is close to or slightly shorter than the average duration in general, while that of phrase-final syllable is considerably longer than the average in general. On the contrary, however, in the phrase that to be completed in sense, the duration of first syllable is longer than that of average, and that of last syllable is close to or slightly shorter than that of average. In addition, such kind of speech tempo adjustment is identified both for News' speech and prose declaim. Consequently, it may reflect a common rule in Chinese. Perhaps, it is another factor to cause the perceptual impression of so-called "isochrony", since such tempo adjustment seems to be relationally invariant, it reoccurs regularly in particular speech position. In the fact, similar phenomenon has also found in stress-timed languages like English (Laver, 1994).

Table IV Distribution of syllable duration in phrase-initial and phrase-final: the phrase (1)to be continued; (2)to be completed

Speaker	Average in general		Average in Phrase-Initial syllables		Average in phrase-final syllables	
	Mean	Sd.	Mean	Sd.	Mean	Sd.
Female	179	60	(1) 168	47	(1) 298	61
			(2) 190	56	(2) 177	33
Male	155	59	(1) 154	53	(1) 250	106
			(2) 219	51	(2) 142	26

4. CONCLUSION

According to the investigations respected to read sentences, New's speech and declaim speech, the duration of syllables, the durational distribution between rhythmic units and the tempo adjustment within certain units are all varied due to linguistic and paralinguistic constraints. Consequently, I would suggest that (1) there is no evidence to regard Chinese as a syllable-timed language; (2) speech rhythm seems more related to regular occurrence and variation of certain prosodic phenomena at particular positions, in stead of the occurrence in equal time interval of syllables, stresses or any other speech constituents.

MAIN REFFERENCES

Abercrombie, D. (1967). Elements of General Phonetics, Edinburgh University Press.

Boomer, D. S. (1978). The phonemic clause: speech unit in human communication(quoted from Laver, 1994).

Cao, Jianfen (1991). Durational patterns of syllables in Standard Chinese. The Proc. of 12th ICPhS, Aix-en-Province, France, August 19-24, 1991

_____(1994).The effects of accentual focus and lexical stress upon temporal distribution in a sentence. The Proc. of ICSLP' 94, Yokohama, Japan, Sept. 18-22, 1994;

_____(1995). Basic temporal structure of a sentence in Standard Chinese. Journal of Chinese Linguistics, Vol. 7, 1995.

_____(1999). Acoustic-phonetic characteristics on the rhythm of Standard Chinese, Proc. of 4th National Conference on Modern Phonetics, Beijing, August 25-27, 1999.

Dittmann, A. T. And L. G. Llewellyn (1967). The phonemic clause as a unit of speech decoding(quoted from Laver, 1994).

Laver, John (1994). Principles Of Phonetics, Cambridge Press, New York.

Miller, G. (1956), The marginal seven, plus or minus two? Psychological Review, 1956.

Passy, Paul (1930). Outlines of Comparative Phonetics, Chinese version Translated by Liu Fo, The Commercial Press, LTD., Shanghai, 1930.

Pike,K.L.(1946). The intonation of American English, 2nd edn, Ann Arbor, University of Michigan Press.