THE RHYTHM OF MANDARIN CHINESE

Cao Jianfen

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

ABSTRACT

This study is concerned with the rhythm of Mandarin Chinese. As the basis of the study, a set of speech materials were selected from TV news and broadcasting. Pitch and duration measurements were made through their spectrograms, and an informal perception test on rhythm unit division was conducted as well. This paper reports some preliminary results obtained here, and the description is concentrated on rhythmic structure, including the division of rhythmic chunks, the hierarchical organization, the coherent features within rhythmic units and the boundary markers between these units. In addition, some related issues are also discussed in general.

1. INTRODUCTION

1.1 General conception

As a part of the prosody of a language, speech rhythm is not easy to define, because it usually interacts with other prosodic constituents in speech flow. One can define it from different points of view. Consequently, the existing definitions of rhythm are manifold and controversial.

Functionally, rhythmic organization is a chunking strategy concerned with both speech production and perception. Modern cognitive psychology suggests that language understanding responds to discourse chunks. Actually, it is also true in human speech perception and speech production. On the one hand, during speaking, especially when uttering a longer sentence or syntactic phrase, people neither utter without any break nor word by word separately, but combine two or more words into a larger rhythmic chunk according to temporary needs of semantic expression. On the other hand, in perception, people are also sensitive to such kind of chunks, instead of individual words. Some studies have noticed that this phenomenon is based on human cognitive mechanism, "the cognitive process of planning speech is typically episodic. That is, an episode of speech is cognitively planned as an unitary entity, and uttered and perceived as an integrated act" (Laver, 1994). Therefore, this is a universal phenomenon existing commonly in human languages, and also a hot issue to be discussed in phonetics, phonology and speech technology.

In consequence, from this point of view, rhythm of a language is essentially referred to the combination of smaller units into larger units during speaking, as well as to the division of utterance into prosodic groups during speech perception or language understanding. The key point under discussion is that what the principle of such combination or division is and how it is made possible.

1.2 Rhythm studies in Chinese

In respect of Chinese rhythm, many contributions have made, but the conception of rhythm was different from one another among these contributions, this situation may be related to the difference of research angles. We can take some examples from the Mainland in the last ten years. Some of them studied from the point of junctures (e.g., Xu, 1986); some of them took viewpoint of aesthetic perception of the text (e.g., Wu, 1992); some research integrated rhythm with stress and intonation (e.g., Shen, 1994); and some of them studied from the angle of syntactic structure (e.g., Wen, 1994; Zhang, 1998). At the same time, there were also several approaches more or less related to acoustic correlates of the rhythm. For example, the studies on the pause distribution in speech flow (e.g., Ye, 1996; Hua, 1998); the study on the relationship between pausing and syntactic structure (e.g., Mao, 1994); and on the prosodic cues to syntactic boundaries (e.g., Yang, 1997). In addition, more and more researchers coming from the field of speech technology have joined to the study on Chinese rhythm by integrating it with intonation or phrasing in corresponding modeling (e.g., Chu etc., 1995; Li, 1998). All of these contributions have made possible to further the study of Chinese rhythm. However, there is still lacking of a general idea on Chinese rhythm, as some foreign scholars complained that "the rules which underline them (i.e., Chinese rhythm) are neither sufficiently rendered in common textbooks, nor do the dictionaries usually offer..." (Triskova and Lomova, 1999).

During this period, the author of this paper has also conducted several investigations that tried to examine this issue from the nature of speech timing (e.g., Cao, 1991;

1992; 1994; 1995a; 1998; 1999). The observations lead to belief that rhythm is objectively referred to temporal distribution of linguistic information of a language, and it can reflect the strength of relationship between speech units on different levels.

Bearing this idea in mind, this paper will discuss rhythm by integrating it with corresponding temporal and spectral manifestations. The main attention is first paid to the description on the division of rhythmic chunks and the hierarchical organization, then, to the examination of the coherent features within rhythmic units and boundary markers between these units. After that, a general discussion and consideration will be conducted. And finally, a brief summary will be given.

2. TEST MATERIALS AND EXPERIMENTAL METHODS

In most of the previous studies, the approaches on Chinese rhythm were based on some designed phrases or sentences. To further observe the situation in natural speech, we try to use discourse as test material.

In this study, two groups of test materials were employed. The first group was selected from TV news announced by one male and one female speaker (hereafter, news speech). Such style of speech is a kind of typical declarative speech, usually produced without special accent and speech mood. Therefore, it is more suitable for a preliminary observation of the basic rhythmic structure. The second group is a set of prose declaims played by a male speaker (hereafter, declaim speech), which was recorded from broadcasting. This style of speech usually has relatively slow tempo and more often rhythmic break, through which, we can further observe the basic characteristics of Chinese rhythm.

Usually, rhythmic division is not easy to be determined according to syntactic parsing within a complex phrase or sentence, and the result may be varied dependently. It can be illustrated by a well-known example on the grouping strategy of “下雨天留客天留人不留”, in which, there is at least two possible ways to make rhythm division:

(1) “下雨天-留客天-留人不留?”
(“The reining day is a day that suitable to ask the guest to stay. Do you ask?”)

(2) “下雨-天留客-天留-人不留.”
(“It’s reining, it is the heaven to ask the guest to stay. The heaven does, but the host does not.”)

Obviously, these two different strategies express quite different semantic contents, while the listener can follow the speaker in either case. This example indicates that different strategies on rhythm division must have different acoustic-phonetic correlates, which form the material base that made possible of explicit communication between speakers and listener. Therefore, we would employ these correlates as a criterion for rhythmic division in this study.

To observe the acoustic-phonetic manifestation of these correlates, pitch and duration measurements were made through related spectrograms, and an informal perception test was conducted as well. For the convenience of description, we will take some examples to specify the results.

3. RESULTS AND DISCUSSION

3.1 The division of rhythmic units

According to relevant studies in respect to different languages, perceived rhythmic break usually corresponds to certain acoustic–phonetic correlates. For example, in Danish, Hansen etc. (1994) found that the rhythmic break within a sentence is usually either carried out by a moderate lengthening followed by a silent pause in the last syllable of the unit, or by a marked lengthening but without silent pause in this case, while the boundary between sentences is always accompanied with a silent interval but no lengthening occurred in this position. In Chinese, a previous study conducted by Xu (1986) also found lengthening phenomenon taken place at rhythmic border, but he did not mention the distinction in different speech levels, which may be due to the speech materials that he tested were limited in smaller pieces.

In the present study, preliminary results obtained from news speech show a situation similar to that of Danish. Both quantitative change in pre-boundary syllable and silent interval change at the border are existed corresponding to rhythmic grouping, and these acoustic-phonetic data do form as a hierarchy which representing different speech levels.

In this study, the division for news speech was first made according to a perception test, in which, the listeners were asked to give out perceived breaks and their strength, and then, the durational data for relevant events were measured and calculated.

The results for two example paragraphs are listed in Table I. In which, each numbered piece corresponds to a perceived intermediate rhythm unit, and the millisecond amount in each bracket is the duration of silent pause measured at that position, and the numerals after each syllables represent durational ratio of corresponding syllable rhyme, which was calculated from measured duration vs. average duration. If the value is higher than 1.00, we consider it signaling a lengthening, and vice versa. All these data objectively reflect temporal manifestation of the news speech. The details will be discussed later.

Table I. Examples of rhythm division and corresponding durational data in the news speech: (a) for female utterance, (b) for male utterance

(a)

(1)中 0.63 央 1.37 军 0.83 委 1.06 主 0.80 席 1.03 江 1.03 泽 0.52 民 2.15 (G:134.4ms)

(2)日 0.80 前 1.35 签 0.88 署 0.60 命 1.01 令 0.96 (G:459.4ms)

(3)颁 1.26 布 0.85 实 0.61 行 1.86 (G:68.8ms)

(4)中 1.03 国 0.89 人 0.67 民 0.73 解 0.67 放 1.12 军 1.56 (G:130.6ms)

(5)军 1.18 事 1.18 交 1.03 通 1.14 运 0.97 输 0.53 条 1.04 例 0.63 (G:606.3ms)

(6)这 0.60 是 0.71 我 0.99 军 1.18 第 0.64 一 0.88 部 1.75 (G:37.5ms)

(7)全 1.19 面 1.39 规 1.16 范 1.67

(8)军 1.04 事 0.72 交 0.91 通 0.92 运 1.15 输 0.74 工 1.02 作 0.67 的 0.79
基 0.68 本 0.88 法 1.11 规 1.18 (G:546.8ms)

(9)它 0.55 的 1.41 颁 1.31 布 0.69 实 0.63 行 1.60 (G:253.1ms)

(10)标 1.35 志 0.75 着 1.06 我 1.38 军 1.28 的 1.73

(11)军 1.07 事 0.60 交 1.00 通 0.95 运 1.14 输 0.58 工 1.03 作 1.16 (G:300ms)

(12)进 0.97 入 0.71 了 0.91 法 1.32 制 0.79 化 1.15 管 1.35 理 0.65 的 0.60
新 0.68 阶 0.67 段 0.93 (G:1112ms)

(b)

(1)红 1.25 塔 0.67 人 1.05 扶 1.00 贫 1.90(G:178ms)

(2)走 1.07 的 0.52 是 1.42 开 1.58 发 1.58 扶 0.80 贫 0.80 的 0.45 路 0.87 子 0.00 (G:719ms)

(3)他 0.39 们 0.66 说 1.08(G:528ms)

(4)不 0.56 能 0.81 只 1.36 是 1.47 冬 1.29 送 1.29 寒 1.62 衣 1.93 春 1.05 送 1.05 粮 1.66(G:484ms)

(5)而 0.72 要 0.83 从 1.18 生 1.35 活 1.36 科 1.12 技 0.81 和 0.72 农 0.72 业 0.97 基 0.97 础 0.00
设 1.01 施 0.89 上 0.52 下 0.56 功 0.85 夫 0.85(G:2000ms)

According to the data shown in Table I, both the timing behavior of each speech unit and the pausing distribution at each rhythmic boundary are quite well matched to the perceived prosodic grouping. Therefore, the divisions for declaim speech in this study were made directly with reference to measured acoustic- phonetic parameters. An example of results is shown in Table II, where the unit-final (i.e., pre-boundary) lengthening and silent interval between units are marked by symbol "-" and "*" respectively. If both of them occurred at the same border, then marked by "-*".

Table II. An example of rhythm division in the declaim utterance:
- pre-boundary lengthening; * silent pause; -* pre-boundary lengthening plus silent pause

在人的一生当中-* 不管你* 爱好酒- 还是讨厌酒-* 或抱- 无所谓的态度* 人们- 或多或少-*
直接- 或间接地* 都在和酒- 搭上关系-* 它留驻在-* 物质生活- 和精神生活的- 各个领域-*
成为一种- 世界性的- 文化现象* ....
客- 从远方来-* 无酒-* 不足以表达- 款款厚意-* 朋- 到远方去-* 无酒- 不足以表示-*
依依深情-* ....

A post hoc perception test reveals that the rhythmic divisions made in Table II are quite well matched with the perceived impression.

In addition, from spectral analysis, we find that the behavior of pitch movement also provides important information for rhythmic division. Fig. 1 shows an example of pitch movement of a sentence from the news speech. The doted curves from top to bottom represent pitch contours for intermediate rhythmic chunks of "它的颁布实行", "标志着我军的", "军事交通运输工作"and "进入了法制化管理的新阶段" respectively. We will discuss these characteristics later in this paper.

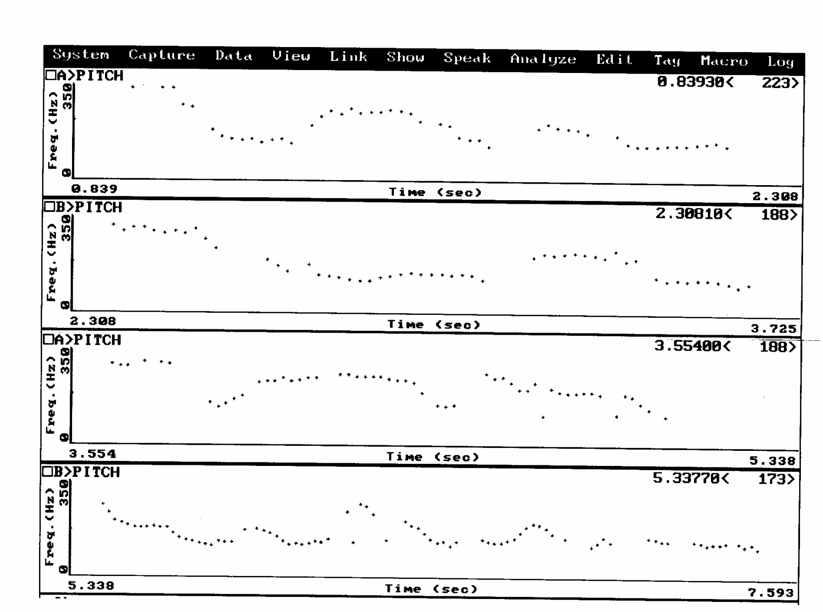


Fig. 1 The history of pitch movement of sentence “它的颁布实行，标志着我军的军事交通运输工作进入了法制化管理的新阶段。”

3.2 Hierarchical organization

According to the results of perceptual judgments and corresponding acoustic-phonetic measurements, we find that the rhythm of Chinese obviously forms as a hierarchy. It consists of three main layers, namely, minor rhythmic unit, intermediate rhythmic chunk and major rhythmic group. Both spectral and temporal analyses conducted here find that rhythmic unit is also a synonym for phonological unit, because it often acts as a carrier of the information on phonological process or intonation. Therefore, for the convenience of discussion, hereafter we employ the prosodic word, prosodic phrase and intonation phrase as their equivalents.

3.2.1 Prosodic word—the minor rhythmic unit

Generally, perceived rhythmic chunk is more related to the higher, instead of lower, speech levels, so the division on low level is not easy to be determined by perception test, especially in relatively speed speech. In this case, related acoustic-phonetic data can provide objective evidence. For example, from Table II, we can find that the minor rhythmic unit usually consists of two or three syllables. In some particular cases, it also contains a few monosyllabic words, such as the “客”and “朋”in Table II, where these monosyllabic chunks are strongly lengthened so that to match with the disyllabic or trisyllabic chunks in rhythm.

As the basic unit, prosodic word is the principal building block of Chinese rhythmic structure. It can be further grouped into larger unit in higher rhythmic layers, or directly play the role of the units in those layers.

Generally, the relationship between syllables in prosodic word is tightly cohered, in fact, prosodic word is the right domain for some phonological processes, for example, tone sandhi and lexical temporal distribution is usually taken place in this size. Consequently, this basic unit is roughly equivalent to the standard foot and super-foot in metrical phonology.

3.2.2 Prosodic phrase—the intermediate rhythmic chunk

Prosodic phrase is the intermediate rhythmic chunk worked between prosodic word and intonation phrase. In some literature, it is also called phonemic clause or phonological phrase. As mentioned above, the cognitive process of planning and perception of speech are both based on a chunking strategy, in which phonemic clauses are processed as successive coherent package of sound, syntax and sense (Boomer, 1978). However, prosodic phrase is difficult to be defined precisely, though it is the most common and important rhythmic unit in speech.

The key point is that how to define its work domain? What is the characterized phonetic property? In respect to Mandarin Chinese, we will discuss this issue more specifically in 4.1.1.

3.2.3 Intonation phrase—the major rhythmic group

Intonation phrase contains two or more prosodic phrases, it is usually identified to syntactically defined sentence. For example, in Table I, the prosodic phrase (1) and (2) combine into an intonation phrase, and (3), (4) and (5) combine into another intonation phrase. From Table II, we can see that in the declaim speech, due to its relatively slow tempo, the intonation phrase tends to contain even more intermediate units. It indicates that the slower the speech rate, the more the rhythmic division.

3.3 Coherent characteristics within rhythmic unit and boundary makers between the units

Rhythm phenomenon reflects temporal distribution of linguistic information and the strength of the relationship between speech units. Generally, the relationship is much tighter when two or more units cohering into the same rhythmic group, and it is relatively loose when they are separated into different rhythmic group. The acoustic-phonetic data obtained in this study do provide strong evidence to signal not only a demarcative cue, but also a coherent signal.

3.3.1 Coherent characteristics within rhythmic unit

In the lowest layer, i.e., within the prosodic words, the coherency is generally represented by relationally invariant patterns for both of tone sandhi and durational distribution. From the pitch contours shown in Fig. 1, combined with the durational data listed in Table I, we can see that either tonal or durational patterns for each disyllabic or trisyllabic chunks are basically identified with their lexical forms (Wu, 1992; Cao, 1989; 1992; 1995a; 1995b). For example, the pitch contour of “管理(的)” presents a typical tone sandhi pattern of so-called “上上相连，第一上变阳平”; and the time relations between syllables in prosodic word, such as that of “交通”and “管理的”, typically show the lexical durational patterns of regular type and neutral type words respectively (Cao, 1989; 1992).

In the intermediate and major rhythmic layers, i.e., within the prosodic phrase and intonation phrase, the coherent property can be observed from two aspects.

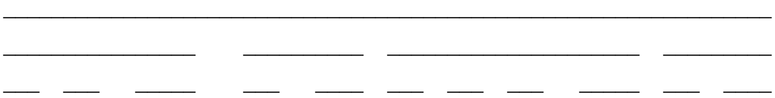
Firstly, the behavior of pitch movement is mainly characterized by an uninterrupted declination tendency of pitch register for the whole unit. For example, in the intonation phrase“它的颁布实行标志着我军的军事交通运输工作进入了法制化管理的新阶段”, the pitch register at the beginning of each sub-units, i.e., of “它” in “它的颁布实行”,“标” in“标志着我军的”,“军” in“军事交通运输工作”,“进” in“进入了法制化管理的新阶段” shown in Fig. 1, are successively lowered without any interruption of this tendency. And similar relation also appears between prosodic words within prosodic phrase if there is no any accentuation occurred in this position. For example, within the prosodic phrase “它的颁布实行”，the pitch register of “它(的)”, “颁(布)” and “实(行)” are successively lowered, too. At the same time, such uninterrupted declination tendency is usually accompanied by a gradual compression of the pitch range from the beginning to the end of a prosodic phrase or intonation phrase. For example, as what shown in the top frame of Fig. 1, the pitch range for the whole prosodic phrase “它的颁布实行” is obviously compressed gradually.

Table III. An example on the correspondence of rhythmic division with acoustic information:
Sht.—shortened; Lnth.—lengthened; Tit.—tight; Los.-/ Los.+ —relatively / obviously loose; FW-function word

Utterance:

这是 我军 第一部 全面 规范 军事 交通 运输 工作的 基本 法规

Rhythmic hierarchy:



Acoustic information:

Sht. Lnth. Lnth.+p. Lnth. Lnth. Sht. Sht. Sht. Sht. Sht.
(+FW)

Rhythmic relation strength:

Tit. Los.- Los.+ Los.- Los.+ Tit. Tit. Tit. Los.- Tit.

Secondly, in temporal aspect, there are also acoustic cues to signal the coherence. It is mainly characterized by a pre-boundary lengthening, instead of silent interval, so that to show relatively tighter relation between sub-units. Table III gives an example of intonation phrase “这是 我军 第一部 全面 规范 军事 交通 运输 工作的 基本 法规”, in which there are totally 11 prosodic words, 4 prosodic phrases. The rhythmic hierarchy and the corresponding acoustic information and relative coherent strength are listed in the frames from top to bottom respectively. From this example, we can see that the strength of rhythmic coherency within prosodic phrase is different from that within intonation phrase. Specifically, in each prosodic phrase, there are usually no or only slightly pre-boundary lengthening taking place at each prosodic word boundary; while in intonation phrase, there are always considerable pre-boundary lengthening occurring in each prosodic phrase boundary, and mostly accompanied by a short silent pause in this case. Of course, there is also strength difference existing at different prosodic word boundaries in the same prosodic phrase. It means that even within a prosodic phrase, some of prosodic words exhibit tighter relationship, such as the situation occurred between “这是” and “我军” or between “军事” and “交通”, “交通” and “运输”, etc.; but some of them exhibit relatively loose relationship, like the situation found between“我军” and “第一部” or between “全面” and “规范”. All these information indicate that the more the shortening at the unit-final (i.e., the terminal syllable of the unit), the tighter the relation between the units, and vice versa, and it is quite well matched to the perceived rhythmic impression.

3.3.2 Boundary markers between rhythmic units

Similar to the situation found in many other languages, pre-boundary lengthening and silent pause, as well as the distinguished behavior of pitch movement, also play a role of demarcation between rhythmic units in Chinese. Generally, different distribution or their combination of these factors signal boundaries on different rhythmic layers, and it exhibits a regular and hierarchical manner.

3.3.2.1 Boundary marker in temporal aspect

As the apparent boundary cue in temporal aspect, there is marked pre-boundary lengthening and /or a silent pause at the boundaries on different layers.

Specifically, on the lowest layer, the boundary marker is characterized by a moderate pre-boundary lengthening, but without silent pause in general. For example, from Table I, we can find such makers between prosodic words of “中央”and “军委”or “红塔人” and “扶贫”, etc. However, this property is quite fragile, it will disappear when the units, which are separated by it, are merged into a larger unit. Hence, in most of the cases, this boundary marker is blot out.

Table IV. Temporal distribution at intermediate and major rhythmic boundary:

A. mean durational ratio(%) of the rhymes in pre-boundary syllables ;

B. mean duration(ms) of the silent pauses at: a. sentence end, b. Paragraph end

A(%)					B(ms)			
Phrase Type Speech Style	Awaited		Ended		Awaited		Ended	
	Mean	Sd.	Mean	Sd.	Mean	Sd.	Mean	Sd.
Female Speaker	1.68	0.31	0.93	0.23	154	97	a.538 b.1112	74 0
Male Speaker	1.55	0.39	0.86	0.0	397	191	a.719 b.2000 +	0 0
News Speech	1.62	0.34	0.90	0.20	276	169	a.629 b.1020	109 628
Declaim Speech	1.33	0.21	1.05	0.28	59	34	a.548 b.2000+	87 0

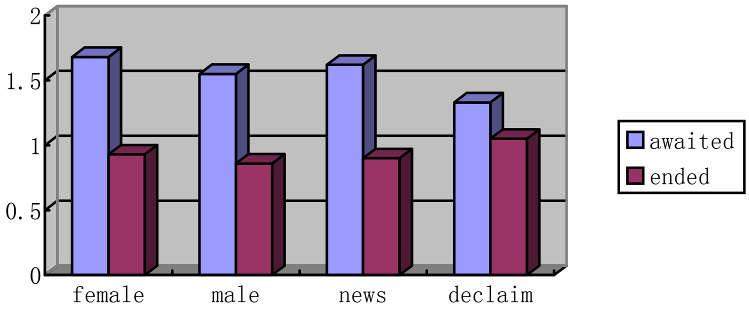


Fig.2 Pre-boundary timing at the boundaries in different style of speech:
awaited—for the units semantically to be continued;
ended—for those semantically to be completed

On the intermediate layer, pre-boundary lengthening is very strong in general, but the specification depends on the unit type of "awaited" or "ended". The situation can be observed from the figures listed in Table IV. Specifically, the figures listed in "awaited" columns represent the boundary marker of prosodic phrase which is semantically to be continued; while those in the "ended "columns represent that of prosodic phrase which is semantically to be completed. Hence it is actually a boundary for intonation phrase. Comparing the figures listed in column 2 and 3 from the left in Table IV, or observing the situations shown in Fig. 2, we can see that the timing situation at the boundary of the "awaited" unit is systematically distinguished from that of the "ended" unit. Specifically, there always exists marked pre-boundary lengthening in the "awaited" case, but does not in the "ended" ones', no matter in what style of speech. It means that this is a regular rule and identically exists in various style of Chinese speech. In addition, as what can be seen from columns 4 and 5 in Table IV, this lengthening phenomenon is usually accompanied with certain amount of silent pause. The duration of its interval is variable, but systematically shorter than that occurring between intonation phrases. For example, according to the data listed in Table I (a), the rhyme of the last syllable of the prosodic phrases (1), (3), (4), (6), (7), (9), (10) and (11) are all considerably lengthened, and often to be accompanied by a short silent pause in most cases.

On the highest layer, i.e., at the boundary between sentences or paragraphs, the first important boundary marker is a strong silent pause, and its interval is systematically longer than that between prosodic phrases. The situation can be observed from the figures listed in the far right column of Table IV. In addition, comparing to the intermediate layer, the pre-boundary timing in this layer shows a quite different picture. As shown in Fig. 2, there is usually no pre-boundary lengthening in this layer, and in most cases, the rhyme in pre-boundary syllable is even shortened. For example, at the end of the phrase"中央...命令" and "它的...新阶段" or "他们说...下功夫", the syllable rhymes are all shortened, in this case, it is always followed by a strong silent pause, typically at the end of a paragraph.

3.3.2.2 Boundary marker in spectral aspect

Resetting of pitch register, i.e., a dislocation of pitch register taken place between rhythmic units is another important boundary marker.

Specifically, the pitch register at the beginning of each unit is higher than that at the ending of the previous unit, this phenomenon can be observed clearly from Fig 1. Usually, this phenomenon occurred between prosodic phrases, like that between "它的颁布实行" and "标志着我军的", is sharper than those between prosodic words, like that between "他的"and "颁布" or "颁布"and "实行"; And this dislocation is even sharper at the boundary between intonation phrases. Resetting of pitch register is a common phenomenon existed in different languages, however, in Mandarin, this resetting has its own characteristic property. The main point is that the resetting in prosodic word level must be restricted by tone features. For example, because the distinctive feature of the 3rd tone is low, therefore, in the case of the first syllable in a unit is a 3rd tone syllable, like the "我" in "我军的" shown in Fig. 1, the pitch register is not always higher but usually lower than that of pre-boundary syllable "着" in "标志着", though here the "着" is a neutral-tone syllable.

4. GENERAL DISCUSSION

4.1What is the relationship between rhythm and the time behavior of speech?

Traditionally, in phonetic theories, rhythm was regarded as a phenomenon related to regular recurrence in time of some given speech unit, and it was suggested that such "isochrony" commonly existed in all spoken languages. Thus, the speech rhythm of different languages is categorized accordingly. For example, it is well known that English is a typical stress-timed language, which means the periodic recurrence of movement is supplied by the stresses-producing process, and French is a typical syllable-timed language, which means syllables recur at a equal intervals of time (e. g., Pike, 1946; Abercrombie, 1967). This view has persisted for centuries in phonetic theories, and is still a very pervasive idea up to date. For example, in a recently published dictionary, rhythm is defined as "the perceptual pattern produced in speech by the occurrence at regular intervals of prominent elements" (Trask, 1996).

In view of this idea, many phonetic studies of rhythm have focused on the search for acoustic evidence for isochrony in order to support the traditional suggestion. However, these studies have shown that the rhythm "has nothing to do with the duration of interstress intervals" (Dauer, 1983), "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" (see Arvaniti, 1994). The results of these investigations indicate that such so-called "isochrony" does not exist even in English and French.

Nevertheless, as a kind of prosodic phenomenon, speech rhythm commonly exists in all spoken languages. People have learned it for a long time, and in perception, it is strongly noticed that rhythm is closely related to the time behavior of speech production and perception. The problem is what is the nature of rhythm? How to view the relationship between rhythmic grouping and speech timing?

In Some earlier studies, rhythm was defined as breath grouping (e.g., Passy 1930). Apparently, this is taking a viewpoint of physiological mechanism, and it may be one source of the idea of the "isochrony", because human breathing is taking roughly in equal time interval. However, at the same time, Passy pointed out that such grouping behavior is also related to sense grouping, i.e., due to the necessary on the turning in semantic expression, in order to provide enough time for perception. Therefore, every breath grouping roughly corresponds to a simple sense unit. Thus, the sound grouping in speech is more or less constrained by logical principle, and such turning is varied, larger or smaller, so the interval of each grouping does not have to be isochrony. Moreover, he specified that breath group can be further divided into group of force, in which there must be one syllable to be accented. Generally, each of such force groupings is combined by two or three syllables. These syllables are cohered very closed in sense, and there is one among them which 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), speech rhythm grouping is not necessary to be isochrony, but seems impossible to be isochrony in most cases.

According to this investigation, we find no base to evidence so-called "isochrony", but do find some interesting phenomena that may contribute to understanding rhythmic grouping. The one is the range of unit length, i.e., the span of syllable number in a rhythm unit; the other is the relationally invariant timing behavior of a rhythmic unit. We will specify the details below in 4.1.1 and 4.1.2 respectively.

4.1.1 The span of syllable number in prosodic phrase

According to the data obtained in this study, a prosodic phrase in Chinese consists of two or more prosodic words. It is usually larger than word but smaller than syntactically defined phrase or clause, and its length is varied all the time. However, the variation seems to have a certain range. If taking an overview to Table I and Table II, we can find that the span of a prosodic phrase is limited to about 7±2 syllables, especially when these syllables occur in relatively unstressed positions. There are only few cases which are beyond this limitation, for example, in the case of "军事交通运输工作的基本法规"shown in Table III, however, we find that in this case, there are always certain function word, such as "的"、 "地" and "和"etc., occurring there to serve as a substitute boundary marker, thus, the span of the sub-division that separated by those function words still follow the general limitation. Besides, comparing the situation shown in Table I to II, we find that the specific span of such intermediate rhythmic chunk is also related to the difference in speech rate and style. For example, according to the data come from the news speech (see Table I), the span is mostly around 7 syllables; while that is mostly around 5 in declaim speech (see Table II).

The length limitation of prosodic phrases observed above may be not surprising, since there are 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, 1968)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, G. 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 to be an accidental event, but a further evidence for the findings made by Dittmann etc. And it may imply a relational invariance related to a common rule on timing control in speech production and perception, hence causing a perceptual impression of appeared “isochrony”.

4.1.2 Temporal structure of prosodic phrases

Table V shows the situation on durational distribution of phrase-initial syllables and phrase-final syllables, in which, “awaited” represents the type of phrase to be continued in sense; and “ended” the type to be completed in sense.

Table V Distribution of syllable duration in phrase-initial and phrase-final: (1)awaited phrase; (2)ended phrase

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

From the data listed in this table, two things related to the temporal structure of prosodic phrases can be seen. First, the duration of phrase-initial syllable is systematically different from that of phrase-final syllable; second, this difference is also clearly conditioned by the type of the phrase. Specifically, in the type of awaited phrase, the duration of the first syllable is close to or slightly shorter than the average duration in general, while that of the last syllable is considerable longer than the average duration in general. However, in the ended phrase, the duration of the first syllable is obviously longer than that in general, and the last syllable of the phrase is close to or slightly shorter than that in general. In addition, this systematic difference described above is identified both for male speech and female speech. Consequently, it may reflect a common rule in Mandarin Chinese. In fact, similar phenomenon has also been found in some other languages including English. For example, Laver (1994:532) reported that both the beginning and ending of utterances exhibit adjustment in speech tempo.

The situations specified above indicate again that speech rhythm has no base related with isochrony, which is at least true in Mandarin Chinese.

In summary, all those described in 4.1.1 and 4.1.2 tell us that speech rhythm may be defined as the regular pattern (or rule) on the temporal distribution of linguistic information, and that such kind of rule is usually manifested as regular occurrence and variation of certain prosodic phenomena at particular positions, instead of isochronic occurrence of any given speech units. This pattern is not only a universal phenomenon existed in different languages, but also characterized language –specifically.

4.2 Can we deduce prosodic structure from syntactic structure?

What is the relationship between prosody and syntax? Can prosodic structure be deduced from syntactic structure? It is a hot issue discussed in linguistics for a long time, and recalling more and more interest and attention of linguists, especially of phoneticians and speech engineers in now days.

The general situation is that rhythmic break must take place at syntactically available boundary, and some approaches in speech technology (e. g., Ozeki, 1997) have found that prosodic information, especially the pause duration can significantly improve the accuracy of syntactic parsing. Based on this situation**,** some researchers suggested that prosodic structure can be deduced from syntactic tree, and many efforts even have been made to detect prosodic structure based on the syntactic structure through some algorithm(e.g., Grosjean et al., 1979; Cooper & Cooper, 1980; Gee & Grosjean, 1983). In these approaches, performance structure, i. e., a kind of structure obtained from experimental data such as pause durations, transitional error probabilities and parsing values in speech, is regarded as quite simply a reflection of prosodic structures. However, the unfortunate fact is that this kind of algorithm not yet has any successful applications up to date, some issues are still left in open, and further study is needed.

Actually, early in 1977, Liberman and Prince have proposed that intonation, rhythm and pausing patterns cannot be directly interpreted by syntactic structure. After that, many other studies related to different languages have found that the identification between prosodic unit and syntactic unit is true only in higher speech levels, but not in lower levels. In consequence, as Rossi (1997) pointed out that syntactic information could be retrievable only partially from prosodic information. Therefore, more and recent studies have recognized that prosodic organization of a language does closely related to its syntactic structure, but far not identified with each other.

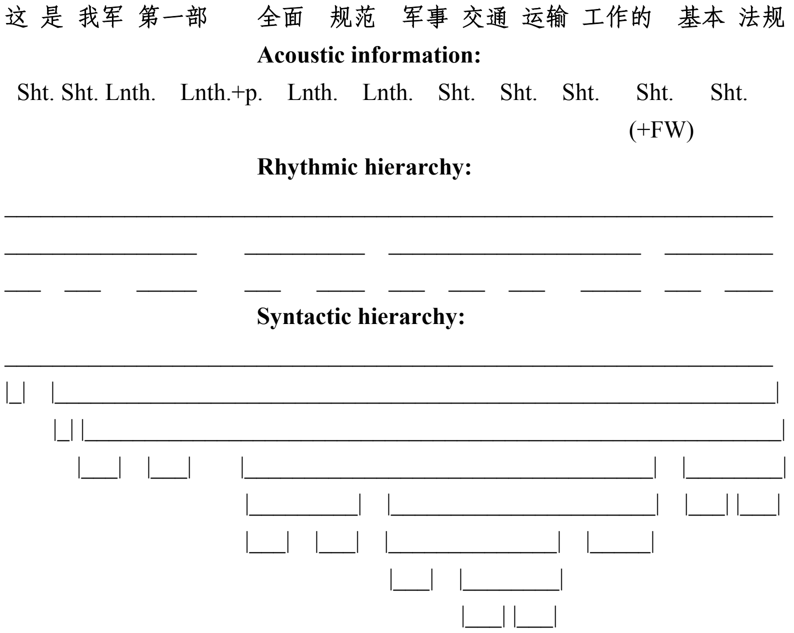

这 是 我军 第一部 全面 规范 军事 交通 运输 工作的 基本 法规
Acoustic information:
Sht. Sht. Lnth. Lnth.+p. Lnth. Lnth. Sht. Sht. Sht. Sht. Sht.
(+FW)
Rhythmic hierarchy:
Syntactic hierarchy:

Fig. 3 A diagram on rhythmic and syntactic organization: sht.—shortening; lnth.—lengthening; Lnth. + p.---lengthening plus silent pause; FW— function word

In the present study, we only give a glance at this point. Fig. 3 shows a typical example in respect to this aspect.

Apparently, the picture of prosodic structure is sharply different from that of syntactic structure. If we take the viewpoint of syntax, the less the branch of the tree (i.e., the higher the level), the strong the boundary strength, and the laxer the relationship between the units, thus, the more the opportunity of pause occurrence at the boundary. Hence, the longest pause in this example should occur between “这” and “是我军……的基本法规”; However, in fact, from the rhythmic hierarchy, we can see that the main rhythmic break occurs in a lower syntactic level, i.e., occurs between “这是……第一部” and “全面…的基本法规”, where it is marked by a pre-boundary lengthening plus a silent pause, it is the strongest break taken place within the sentence. Usually, this kind of conflicting phenomenon between syntax and prosody exists typically in unbalanced sentences, like this example. Only when the sentence is balanced, i.e., only in the case of syllable numbers of each direct constituent in a sentence is roughly equal to each other, the structures for syntax and prosody may be roughly identical. However, due to the limitation on test materials employed in this study, we cannot precisely evaluate the weight of balanced and unbalanced sentences in natural speech. In spite of this limitation, our data obtained here at least provide some further evidences to indicate that we can not wish to derive prosodic organization completely from syntactic structure, and vice versa.

4.3 How to view the connection and demarcation between rhythmic units

A rhythmic unit in certain speech layer usually consists of several sub-units, on the one hand, within this unit, those sub-units form a connective relationship, on the other hand, as independent components of the unit, those sub-constituents are demarcated from each other. And it is true not only from the viewpoint of speaking, but also speech perception. Obviously, it is the same phonetic reality related to different rhythmic events. Then, how to view this phenomenon? According to the present investigation, we may understand it from some material base.

As the objective material base, pre-boundary (or unit-final) lengthening, silent pause and special behavior of pitch movement are the main correlates referred to speech rhythm. These parameters have found commonly in many languages including Mandarin Chinese. From the results obtained in the present investigation, we find that the role of the above acoustic-phonetic parameters is multiple, and their function not only complement to each other, but also restrict to one another.

First, different behaviors of the same parameters may signal different rhythmic phenomena. For example, usually, the un-interruption of the declination tendency of pitch register in a prosodic phrase or intonation phrase has a connective function signaling coherence within the unit, while a break, i.e., resetting in the declination tendency has a demarcative function signaling a unit boundary.

Second, certain combination of different parameters may reflect different rhythmic feature. For example, a pre-boundary lengthening is usually more related to connection relationship between rhythmic units, at the same time, it also imply the existence of the unit boundary, especially when it is accompanied with a silent pause.

According to the situation observed in this study, the lengthening phenomenon usually takes place in the last syllable of prosodic words or prosodic phrases, and it is more prominent in the later case, but seldom occurring in that position of intonation phrases. This situation does match to the difference on rhythmic distance in perception between these layers, i.e., it corresponds to such a reality: the coherence between prosodic words is much tighter than that between intonation phrases. However, you may noticed that lengthening occurred at the end of a prosodic phrase is more prominent than that at the end of a prosodic word, it does not seem to match with their distance difference. To understand this phenomenon, let us see another relevant aspect, i.e., the distribution of silent pause.

Generally, silent pause is the main feature to cue the demarcation between rhythmic units, and the longer the pause interval, the stronger the rhythmic boundary is. According to the data listed in Table I or Table IV, the pause interval does form a hierarchy. Specifically, there is no silent pause between prosodic words within a prosodic phrase, while a relative short silent pause usually existed between prosodic phrases within an intonation phrase, and there is always a longer silent pause occurring at the end of intonation phrase, especially at the end of paragraphs, where the duration of silent pause is around 2000ms. Apparently, this situation is well matched with the exact hierarchy of rhythmic boundary strength. However, at the same time, the quantitative difference existed between these pauses also play the role of connection between units by combining with certain distribution of pre-boundary lengthening. Therefore, a longest pre-boundary lengthening occurring at the end of a prosodic phrase is not surprising, because there is a relative short pause there at the same time. Such kind of combination typically reflects the special status of this intermediate unit in speech rhythm. That is, on one hand, prosodic phrase is a most important and common division related to perceived rhythm, hence, it is usually marked by a short silent pause; on the other hand, however, prosodic phrase is a connective bridge worked between prosodic word and intonation phrase, so it must be marked by a significant pre-boundary lengthening, so as to show more connection, instead of demarcation, in this position.

The acoustic-phonetic manifestation described above indicates that rhythmic connection and demarcation is actually a relative contrast, they form the two organic aspects of the juncture between speech units.

5. SUMMARY

Based on the preliminary investigation conducted here, I would make some suggestion as follows.

(1)Rhythmic structure of Mandarin Chinese contains three basic layers, namely, prosodic word, prosodic phrase and intonation phrase. Each of them is characterized by particular prosodic coherency and boundary features. These features are manifested by certain acoustic-phonetic parameters both in temporal and spectral aspects, **and** they work in a hierarchical manner: in the low layer, the acoustic effects is a moderate pre-boundary lengthening without any silent pause; while in the high layer, is always a sharp silent pause with a pitch resetting, and usually without pre-boundary lengthening; and in the intermediate layer, it is usually carried out by a prominent pre-boundary lengthening together with a short silent pause, as well as a resetting of pitch register. It is clear that both special and temporal cues are combined to signal the coherence and boundary hierarchy of rhythmic structure.

(2) Speech rhythm can reflect the strength on time relation between linguistic units. It may be more related to regular occurrence and variation of certain prosodic phenomena

at particular positions, instead of isochrony of prosodic unit.

(3) Generally, rhythmic break must take place at syntactically available boundary, in this sense, and only in this sense, we can say that prosodically defined unit is identified with syntactically defined unit. However, there is no converse theorem in this relation. Therefore, to wish to detect prosodic structure based on syntactic structure is only a fine belief or ideal, rather than a reality.

The information described above may be of benefit to the automatic segmentation and transcription in speech processing, as well as language teaching. However, this is only preliminary study to very limited materials, and has not considered the possible effects coming from stress, intonation and so on. Therefore, more intensive investigation is needed.

REFERENCES

ABERCROMBIE, D. 1967. Elements of General Phonetics, Edinburgh University Press. ARVANITI, Amalia. 1994. Acoustic features of Greek rhythmic structure, Journal of Phonetics, 22.

BOOMER, D. S. 1978. The phonemic clause: speech unit in human communication, in A. W. Sigman etc., (eds) Nonverba Behavior and Communication, Hillsdale NJ: Lawrence Erlbaum Associates(quoted from Laver, 1994).

CAO, Jianfen. 1990; 1991. Durational patterns of syllables in Standard Chinese. The Proc. of 12th ICPhS, Aix-en-Province, France, August 19-24, 1991; The original work appears in RPR-IL(CASS)/1990[the Report of Phonetic Research, Institute of Linguistics, Chinese Academy of Social Sciences].

__________. 1989; 1992.Temporal structure in bisyllabic word frame: An evidence for relational invariance and variability from Standard Chinese. Proc. of ICSLP'92, Alberta, Canada, Oct. 12-16, 1992; The original work appears in RPR-IL(CASS).

__________. 1992-1993; 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; The original work appears in RPR IL(CASS).

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

__________. 1995. Tone sandhi and stress contrast. Zhongguo Yuwen, No.4.

__________. 1998. A preliminary study on the rhythm in Mandarin Chinese. RPR- IL(CASS).

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

COOPER, W. & Paccia-COOPER. 1980. Syntax and Speech, Cambridge, MA: Havard Univ. Press.

CHU, Min and lU, Shinan. 1995. High intelligibility and naturalness Chinese TTS system and prosodic rules, Proc. of XIII ICPhS, Stockholm, Sweden.

DAUER, R. M. 1983. Stress-timing and syllable-timing re-analyzed, Journal of Phonetics, 11: 51-62.

DITTMANN, A. T. & L. G. LLEWELLYN. 1967. The phonemic clause as a unit of speech decoding. Journal of Personality and Social Psychology 6(quoted from Laver, 1994).

_____________________________________. 1968. Relationship between vocalization and head nods as listener response, Journal of Personality and Social Psychology 9(ibid).

GEE, J. P. & GROSJEAN, F. 1983. Performance structures: psycholinguistic and linguistic appraisal, Cognitive Psychology, 1983, 15:411-458.

GROSJEAN, F., GROSJEAN, L. & LANE, H. 1979. The patterns of silence: Performance structures in sentence production, Cognitive Psychology, 1979, 11:58-81.

HANSEN, P. M. etc. 1994. Syntax, pauses and temporal relations in the final part of the sentence, Working Papers, Linguistics, Lund University, Vol.41.

HUA, Wu. 1998. A perceptual experiment on the distribution of pause in read speech, The Proc. of the Conference on Phonetics of the Languages in China, City University of Hong Kong, May 28-30.

KOHNO, M & TSU, Shima. 1989. Rhythmic phenomenon in a child's babbling and one-word sentence. The Bulletin No. 191, The Phonetic Society of Japan.

LAVER, John. 1994. Principles Of Phonetics, Cambridge Press, New York.

LI, Aijun. 1998. Durational characteristics of the prosodic phrase in Standard Chinese, The Proc. of the Conference on Phonetics of the Languages in China, City University of Hong Kong, May 28-30.

LIBERMAN, M. & PRINCE, A. 1977. On stress and linguistic rhythm, Linguistic Inquiry, 8.

MAO, Shizhen. 1994. A preliminary study on grammatical pause in Modern Chinese, Working Papers of Huadong Normal University, No. 2.

MILLER, G. 1956, The marginal seven, plus or minus two? Psychological Review.

OZEKI, K., KOUSAKA, K. & ZHANG, Yujie. 1997. Syntactic information contained in prosodic features of

Japanese utterances, Proceedings of ESCA, Eurospeech’97.

PASSY, Paul. 1930. Outlines of Comparative Phonetics, Chinese version translated by LIU Fo, The Commercial Press, LTD., Shanghai.

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

ROSSI, M. 1997. Is syntactic structure prosodically retrievable? Proceedings of ESCA, Eurospeech’97.

SHEN, Jiong. 1994. The organization and types of Chinese intonation. Fangyan, No. 3.

TRASK, R. L. 1996. A Dictionary of Phonetics and Phonology. Routledge, London and New York.

TRISKOVA, H. & LOMOVA, O. 1999. Preliminary announcement of International Workshop on Tone, Stress and Rhythm in Spoken Chinese.

WEN, Lian. 1994. Rhythmic aspect of Chinese discourse, Zhongguo Yuwen, No.1.

WU, Jiemin. 1992. The period and hierarchy of Chinese rhythm. Zhongguo Yuwen, No. 2.

WU, Zongji. 1992. An Outline of Modern Chinese Phonetics. Chinese Language Teaching Press, Beijing.

XU, Yi. 1986. Acoustic-phonetic characteristics of the junctures in Standard Chinese. Zhongguo Yuwen, No.5.

YANG, Yufang. 1997. Prosodic cues to syntactic boundaries, Acta Acustica, Vol. 22, No. 5.

YE, Jun. 1996. Acoustic cues of pause, The Proc. of 3rd National Conference on Phonetics in China, Beijing.

ZHANG, Bin. 1998. The Grammar of Chinese. Shanghai Educational Press, Shanghai.

汉语普通话的节奏

曹剑芬

中国社会科学院语言所

摘要

本文研究汉语普通话的节奏问题。研究的主要基础是对电视新闻联播和电台广播话语的实验分析， 包括主观听辨试验和客观的音高和时长测量。根据初步的实验结果，重点介绍和讨论以下几方面的内容:

一、节奏组块的划分;
二、节奏的层次结构;
三、节奏单元内部的内聚特征;
四、节奏单元之间的分界标志;
五、讨论

5.1 节奏同话语信息时域分布的关系;

5.2 节奏结构同句法结构的关系;

5.3 节奏组块的分与合的关系;

六、小结

6.1 汉语普通话的节奏包含韵律词、韵律短语和语调短语三个基本层次; 韵律词通常包含 2-3 个音节，韵律短语的跨度多数为 7±2 个音节。

6.2 节奏单元的内聚特征和分界标志，主要体现为各个语音单元音高的规律性起伏变化和时长的规律性伸缩停延;

6.3 韵律节奏的结构是以句法结构为基础的，但不等于句法结构，因而不能期望完全通过句法结构推导节奏的层次结构。

6.4 语音的节奏看来并不是建立在某种语音成分或语音单元如重音或音节的等间隔出现的基础上，而是建立在语音信息在时间域的规律性分布的基础上，具体表现为一定的韵律现象在一定位置上的规律性出现。这种规律性的出现模式客观上体现了口头话语的层次结构。