SYLLABIC VOLUME AS AN ACOUSTIC CORRELATE OF METRICAL STRUCTURE AND FOCUS IN MANDARIN.

Xie Guanghua

Department of Linguistics
The Australian National University

ABSTRACT - This paper investigates the use of syllabic volume (a three dimensional acoustic value) as the acoustic correlate of metrical structure and focus in Mandarin.

AIM

One of the main assumptions of metrical phonology is that the relative prominence of each constituent in a metrical tree is indicated by nodes at different levels. Each node is labeled either s or w where s means 'stronger than' and w means 'weaker than' (Hogg & McCully 1987: 62-75). However, how can the relationship between a strong and weak constituent be interpreted physically? So far there has been little work to find the acoustic correlate or evidence of relative prominence for each constituent or to quantify how strong or weak a constituent is. In this project I show that the syllabic volume of a constituent 's' is greater than that of a constituent 'w'. Thus strong and weak nodes in the metrical structure are not only assumed theoretically but also supported by measurable values of the relevant constituents.

METHOD

The metrical structure in the sample sentence *jintian ta ban jia* (She/He is moving house today) and focus such as *JINTIAN ta banjia* (She is moving house **today**), *jintian TA banjia* (**She** is moving house today) and *jintian ta BANJIA* (She is **moving house** today) were elicited from 6 native speakers. The sample sentence, and the phonetic structure of each syllable in the sample sentence, is suitable for observing some phenomena with which we are concerned. For instance, this sentence is uttered in a single breath group and it consists of four lexical items. All syllables in the sample sentence have the same high tone and all the initial consonants of the syllable are stops. In order to examine metrical structure and focus placement I also designed a text and a dialogue in which some particular linguistic phenomena occur.

Prosodic features, rhythm and focus in particular, are manifested in three acoustic dimensions. Any single value of dimension i.e. fundamental frequency (Fo), amplitude (Ar), or duration cannot describe rhythm and focus properly. The syllabic volume which is a three dimensional acoustic value is calculated in a way similar to Kratochvil (1969, 1987) for each syllable of the sentence. The hypothesis of syllabic volume is based on the concept of physical feature configuration. The relation between the length, width, and height of an object corresponds to the relation between the duration, amplitude, and fundamental frequency of a sound. The formula which Kratochvil (1969: 9) proposes for calculating the syllabic volume is expressed as: V=Fo*Ar*T/1000. Fo and Ar represent the mean value of fundamental frequency and overall amplitude and T the duration of the given syllable's vowel (or vowel plus final nasal). For simplicity, the order of magnitude is reduced by three steps.

The revised formula for calculating syllabic volume in this project is expressed as: V=[Dc*Arc*100]+[Dv*Arv*Fov]/1000. The additional part of the formula is Dc*Arc*100 where Dc represents the duration of initial consonants excluding sonorants, and Dv indicates duration of vowels plus sonorants. In Mandarin liquids, nasals, and semi-vowels can occur at the beginning of the syllable, but among the consonants only nasals can occur at the end of the syllable. As liquids, nasals, and semi-vowels can be stretched into a distinct column of harmonics on the spectrogram, the duration of these sonorants can be counted with vowels. The original formula does not take into account the value of the initial consonant. The reason, as Kratochvil (1969: 16) explains, is mainly for the sake of simplicity. If only pitch or pitch contour of the syllable is observed, of course we need only investigate the distinct column of harmonics. But if syllabic volume is to be calculated, the whole syllable, including the initial consonant, should somehow be taken into account.

Recording was done on professional equipment in the recording laboratory of the Linguistic Department of the Australian National University. After recording each informant's sounds on high quality tape, I made three spectrograms for each type of the sample sentence of each person. The first one is a wide band spectrogram with a normal linear scale. The second one is a narrow band spectrogram with an expanded linear scale. The third type of spectrogram is an amplitude display in which each space between two horizontal lines indicates 6 decibels (dB).

How the boundaries of each syllable are identified from the spectrogram is the key for measurement. Consonants and vowels may be uttered simultaneously but, as they are realized sequentially, the boundary of syllables always overlaps (Borden & Harris 1980: 42-5). For instance, in *tian* and *ta* the vocal cord impulse which is part of the *tian* will last into the hold phase of the initial consonant of *ta*. The final part of *tian* and the first part of *ta* overlap. That is why we can see the boundary between *ta* and *ban* but we cannot see the point of onset of oral closure between *tian* and *ta* . This is true for careful speech. When speech becomes faster, there is no possibility of finding an identifiable point to reflect the boundary. In order to identify the boundaries of syllables, I put three kinds of spectrograms together. By comparing the harmonic signals from the low frequency scale with the white noise from the high frequency scale, and comparing the wide band spectrogram with the amplitude display, each syllable can be segmented adequately. I used the initial consonant of a syllable as a clue to identify the boundary of syllables. The sentence *jintian ta ban jia* consists of five syllables. The initial consonant of the first and fifth syllable is a voiceless palatal affricate. The second and third are aspirated alveolar stops. Both affricate and aspiration show random energy distributed unevenly throughout the spectrogram. The initial consonant of the fourth syllable is an unaspirated bilabial stop. The plosive is characterized by the acoustic consequences of its pressure and release stages. There is no energy during the pressure period but this is followed by a brief burst of energy at the moment of release. The boundary of each consonant above 3000 Hz can be identified from the spectrogram. The initial consonant of the last syllable is a bilabial nasal which has a clear formant structure which distinguishes it from the vowel /a/.

RESULTS

Metrical Structure

According to the grand mean syllabic volume, the strong-weak or weak-strong rhythmic pattern can be identified at different levels of the sentence. Examining the mean syllabic volume of the four types of the sentence *jintian ta ban jia* (declarative, interrogative, interrogative plus *ma*, and exclamatory plus *ya*) in table 1.we find that, although the individual values in each type of sentence are different, the distribution of syllabic volume in the four types of sentences follows the same basic rhythm pattern. Firstly, based on the grand mean syllabic volume, the six syllables can be divided into four groups which are *jintian*, *ta*, *ban jia*, and *m/ya*. Since the grand mean values of *ban jia*, (59 and 85 respectively), and *jintian*, (45 and 47 respectively), are greater than *ta* (26) and *m/ya* (42), we can work out the rhythmic pattern of the four groups in the sentence as strong (*jintian*), weak(*ta*), strong(*banjia*), weak(*m/ya*). Secondly, the grand mean value for *jin* is similar, although slightly smaller, to that of *tian*, and the value of *ban* is similar, although smaller, to that of *jia*. From the grand mean value we can also identify the rhythmic pattern in *jintian* and *banjia* as weak(*jin*, *ban*)-strong(*tian, jia*).

		JIN	TIAN	TA	BAN	JIA	M/YA
Declarative	X	36	35	25	39	46	
	SD	20	13	7	12	15	
Interrogative	X	50	54	25	70	107	
	SD	26	27	6	20	36	
Interrogative	X	51	51	29	65	87	46
+Ma	SD	34	27	13	30	29	20
Exclamatory	X	42	49	25	63	87	37
+Ya	SD	24	24	11	44	39	21
Grand mean	X	45	47	26	59	82	42

Table 1. Syllabic Volume and Standard Deviation of the Four Types of Sentences: Mean of 6 speakers

The main idea of metrical phonology is that phonological strings are not merely linear sequences but are hierarchically structured. The stress patterns reflect, to some extent at least, relations of prominence between constituents of sentences. From the results I propose that the syllabic volume of a constituent 's' is greater than that of a constituent 'w'. Conversely, the syllabic volume of a constituent 'w' is smaller than the syllabic volume of constituent 's'. Thus the strong and weak constituents in the metrical structure and focus in Mandarin are not only assumed theoretically but also supported by measurable values of the relevant constituents. According to the grand mean syllabic volume of the four types of sentence we can draw a schematic diagram of the hierarchical rhythmic structure as shown in figure 1.

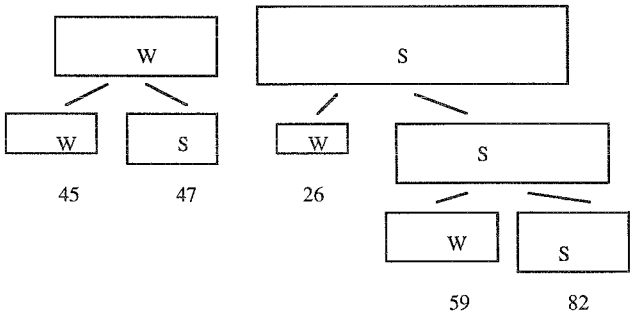

Figure 1 The Hierarchical rhythmic structure of the sentence *jintian ta ban jia*

There are three levels and two kinds of constituents in the hierarchical rhythmic structure of the sentence *jintian ta banjia*. Firstly, the structure is supported by the phonetic facts. The value of each syllable reflects the metrical structure clearly. Secondly the syntactic structure of the sentence is in coordination with the rhythmic structure. At the bottom level, *ban jia* is a verb phrase. In most cases an object is stressed on a verb phrase in Mandarin sentences (Liu, Deng & Liu 1981: 264). A difference between the syntactic and rhythmic structure is that *jintian* is a unit which need not be separated in the syntax, but it has to be segmented in the rhythmic structure. Thirdly, the pragmatic structure of the sentence is reflected in the rhythmic structure as well. *jintian* is the topic at the high level, and *ta* is the topic at a lower level. Comparing with the topic (*jintian* or *ta*) the comment (*ta ban Jia* or *ban jia*) represents the new information which is emphasized. This is the reason why one finds two comments at different levels with both being stronger in the rhythmic structure.

The rhythmic structure of *jintian* requires some comment. *jintian* is a disyllabic noun. The mean syllabic volumes of *jin* and *tian* in the declarative sentence are 36 and 35 respectively. In the interrogative plus *ma* sentence both *jin* and *tian* have the same value 51. In spite of this I still interpret their relation as weak-strong. Why? Actually, the mean values obscure the weak-strong relationship in *jintian*. If we look at the individuals' results, a pattern of weak-strong is clear. Two (DQ, WXL) of the six informants have a w-s pattern in *jintian* for all 4 types of sentences. Another two (LYP, ZSY) have a w-s pattern for 3 of the 4 types of sentences. One (GM) has a w-s pattern for 2 of the 4 types of sentences. There is only one informant (HYH) who has a w-s pattern for only 1 of the 4 types of sentences. Therefore, it is obvious that a weak-strong pattern is the norm for *jintian*. This is in agreement with the acoustic results of Lin, Yan & Sun (1984: 57-73), who found that 70% of disyllabic words in which there is no atonic syllable are produced following a weak-strong pattern. But this means that 30% do not follow this pattern. The syllabic volume of *ta* is the lowest in all four types of sentences. The reason is *ta* is a pronoun. From a pragmatic point of view although *ta* in the sentence is a part of the comment *ta ban jia*, a pronoun generally indicates old information and is pronounced weakly.

Focus

The placement of focus can be understood in terms of the notion of 'information units'. The number of positions where a focus can be placed depends on the number of information units in a sentence. There are three information units in the sample sentence. The first one is *jintian* (today) which denotes when something happens. The second is *ta* (she/he) which indicates the agent. The third is *ban jia* (to move house) which indicates the action. The number of information units in a sentence is decided not only by the syntactic structure, but also by semantic relations.

According to the syllabic volume of the declarative sentence with focus (refer to table 2) I propose to show how the focus in various positions of the sentence can be formalized in rhythmic structure. As the focus which we examined is only in the declarative sentence, not in each type of sentence, the rhythmic structure of the plain declarative sentence should be the basic rhythmic structure for presenting focus. The way focus in the declarative sentence is realized is different from that for rhythm in the four types of sentences. Not only is the value of the constituent with focus greater than that of non-focused ones, but also the rhythmic pattern in the sentence is changed.

		JIN	TIAN	TA	BAN	JIA
Declarative	X	36	35	25	39	46
	SD	20	13	7	12	15
Focus on JINTIAN	X	74	50	19	24	34
	SD	23	21	7	8	16
Focus on TA	X	39	60	57	31	35
	SD	22	31	19	15	23
Focus on BANJIA	X	40	56	24	60	71
	SD	28	38	10	27	29

Table 2. Syllabic Volume and Standard Deviation of Focus in Declarative Sentences: Mean of 6 Speakers

The syllabic volume of the plain declarative sentence is 36, 35, 25, 39, and 46 (see above table). I have explained that the syllabic volumes of *jintian* 36, 35, which are the variations of the grand mean 45, 47, obscure the basic rhythmic pattern in the declarative sentence. Therefore, although syllabic volumes of *jintian* are 36, 35 the basic rhythmic pattern is still weak-strong. When *jintian* is focused the syllabic volumes of *jintian* are 74 and 50 which are greater than those for *ta ban jia* which are 19, 24, and 34. The weak-strong rhythmic pattern of the sentence is changed into strong-weak. The nature of the nodes at the top level of the tree is changed and this affects the nature of the nodes at the low level in the beginning of the sentence but not in the final part of the sentence. This is one of the rules governing constituent change. The rhythmic structure of the sentence can be illustrated as follows:

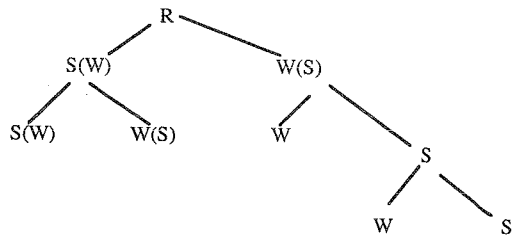

Figure 2. Focus on *jintian* of the sentence
R: rhythm, S: strong node, and W: weak node

When *ta* is focused the syllabic volume of *ta* is 57, which is double than that of non-focused *ta* . However, it is still less than that of *tian* which is 60. The reason is that *ta* is a monosyllabic pronoun whose original syllabic volume is less than that of other words. It also indicates that focus not only affects the word itself but also makes the preceding node stronger. The weak-strong rhythmic pattern at the middle level is changed into strong-weak. When *banjia* is focused the syllabic volumes of *banjia* are 60 and 71 which are bigger than those of other constituents. For the focused element *banjia*, at the middle level the strong node becomes more strong (s+) and at the lower level the weak node becomes less weak (w-) and the strong node becomes stronger (s+). Focus can occur at the beginning, middle, or end of the sentence. There are two ways for focus to be realized in the rhythmic structure. When focus is put on a weak constituent of the sentence, the constituent will become strong. When focus is put on a strong constituent, it will become stronger. The other constituents at the same level still retain their original rhythmic pattern. This is a regular change of rhythmic structure affected by focus.

CONCLUSIONS

What is the main feature of the metrical structure of the sentence *jintian ta ban jia*? After examining the realization of basic rhythmic pattern and focus we can make some conclusions. The whole rhythmic structure (from the top to bottom level) of the sentence *jintian ta ban jia* without focus follows a weak-strong pattern. There are a noun phrase *jintian* plus a clause *ta ban jia* at the top level of the sentence. The second part *ta ban jia* is stronger than the first one. There is a noun phrase *ta* plus verb phrase *ban jia* at the middle level of the sentence. The verb phrase *banjia* is stronger than noun phrase *ta*. There is a verb phrase *ban jia* at bottom level. The noun *jia* is stronger than the verb *ban*. There is a noun *jintian* which consists of two syllables. Since neither syllable is atonic, the second syllable is stronger than the first. In metrical terms we obtain the following formulas:

NP+S $\longrightarrow$ NP+'S/-focus

NP+VP $\longrightarrow$ NP+'VP/-focus

VP $\longrightarrow$ VP+'NP/-focus

S+S $\longrightarrow$ S+'S/-atonic (where S = syllable)

These formulae can be generalized to the single statement:

C+C $\longrightarrow$ C+'C/-focus or atonic (where C = constituent).

We have discussed the realization of the hierarchical rhythmic structure with focus in the sentence *jintian ta ban jia*. The rules governing the constituent changing in the hierarchical rhythmic structure are as follows:

1. S $\longrightarrow$ S+/+focus i.e. a strong constituent with focus will become stronger.

2. W $\longrightarrow$ S/+focus i.e. a weak constituent with focus will become strong.

3. W $\longrightarrow$ S low/W $\longrightarrow$ S high/#_ i.e. when a weak constituent at the high level in the beginning of the sentence becomes strong it will make the weak constituent strong at the low level.

4. S $\longrightarrow$ W low/W $\longrightarrow$ S high/#_ i.e. when a weak constituent at the high level in the beginning of the sentence becomes strong it will make the strong constituent weak at the low level.

5. W(S) $\longrightarrow$ W(S) low/S $\longrightarrow$ W high/_# i.e. when a strong constituent at the high level in the final part of the sentence becomes weak it will not affect the constituent at the lower level.

6.W $\longrightarrow$ W- low/S $\longrightarrow$ S+ high/_#

S $\longrightarrow$ S+ low/S $\longrightarrow$ S+ high/_# i.e. when a strong constituent at the high level in the final part of the sentence becomes stronger, the weak constituent at the lower level will become less weak; the strong constituent will become stronger.

Since the hierarchical rhythmic approach combines hierarchical analysis and quantitative study, we can, based on the acoustic value, predict which constituent is weak or strong at different levels of sentences.

REFERENCES

Borden, Gloria J. & Harris, Katherine S. (1980), *Speech Science Primer-Physiology, Acoustic, and Perception of Speech*, (Willams & Wilkins:Baltimore)

Hogg, Richard & McCully, C,B. (1987) *Metrical Phonology: a Coursebook*, (Cambridge University Press: London)

Kratochvil, P. (1969), "Syllabic Volume as Acoustic Correlate of Prominence in Peking Dialect", Unicorn 5, 1-28

Kratochvil, P. (1987), "Atonicity in Normal Beijing Dialect Speech", Zhongguo Yuwen, 5, 330-45

Lin Maocan, Yan Jingzhu & Sun Guohua (1984), "Preliminary Experiments on the Normal Stress in Beijing Disyllable", Fang Yan:1, 57-73

Liu Xun, Deng Enming & Liu Shehui, (1981), *Practical Chinese Reader,* (The Commercial Press: Beijing)