Intonation Structure of Spoken Chinese: Universality and Characteristics

CAO Jianfen

Abstract

Chinese intonation is extremely complex, and people have been puzzled over decades. Because Chinese is known as a typical tone language, its intonation structure has both universal aspect and special aspect, and the two aspects are both manifested through the tonal features of individual syllables. Thus, a reasonable question may be raised that how the individual syllable can be said in any intonation without losing its lexical identity. Our investigations find that, in the context, tonal variation related to pitch contour of a syllable is relatively limited, but the variation of pitch register is quite free. Therefore, in context, the relative stability of pitch contour ensures the lexical distinction while the variability of the register forms the basis of the simultaneous superposition between local tones and global intonation. Consequently, in this sense, the variability of the syllable register is the key point to understand the relationship between tone and intonation, including accent and rhythm. It serves as a joint, which makes possible of a syllable synchronically carries the information of lexical tones and intonation, and through which, the integration between universal aspect and special aspect of intonation is achieved.

1. Introduction

Intonation in spoken Chinese is a huge topic in the study of prosody. Broadly speaking, it involves all the subjects of the whole prosody. Consequently, it is particularly difficult to be interpreted. However, intonation is mainly characterized by tonal feature, therefore, it can be specified through the variations of tonal feature. The issues related to Chinese intonation involve the generally challenging aspects and some special aspects.

1.1 General Challenge

It is commonly agreed that intonation of speech has multi-layers' function in communication, and each layer has its own unique working pattern and may be integrated with other layers' (*e.g.*, Botinis *et al.*, 2001; Sun, 2002). However, due to the extreme complexity of intonation organization, it still remains a controversial issue that how the different layers of intonation are integrated. The main argument occurred between the superposition model (*e.g.*, Fujisaki, 1982, 1988) and linear model (e. g., Pierrehumbert 1980). For example, Fujisaki's model of intonation consists of accent component and phrase component, and the output of accent component adding to that of phrase ones will be the result of the detailed pitch contour of a sentence. This model

suggested that in phrase layer generally contains an explicit form of entire intonation component, which determines the global trend of the F_0 movement. While Pierrehumbert claimed that larger intonation pattern should be connected by a series of smaller intonation units, and contended that superposition model signifying the global intonation pattern should be pre-planned by the speaker, but such pre-planning seems hard to be believed in cognition. Xu (1999) also argued that there seemed no explicit form of entire intonation. His theory will be briefly introduced later.

1.2 Special challenges related to Chinese intonation

In respect to spoken Chinese, tonal aspects are rather significant in their function, and particularly complex in their variation. Functionally, tonal variation can signal lexical distinction, speech prominence, rhythmic organization as well as the intonation structure. These aspects are synchronically carried through the same entity of fundamental frequency, but behave in different way and are realized as different prosodic component in natural speech, thus resulting in a dazzling F_0 contour in surface. Consequently, how to expose the relationship among these components has become an urgent issue in phonetics theory, and how to decompose them from the surface F_0 contour is one of the most difficult tasks in spoken Chinese processing.

First of all, the challenge is concentrated on what the Chinese intonation is like.

Intonation is the melodic pattern of a language, universally regarded as the F_0 contour of a sentence. For example, in English, it is falling for declarative sentences, and rising for interrogative ones. Thus, with respect to Chinese intonation, people usually just follow this view in the early stage. However, with the deepening of the approaches in Chinese phonetics, more and more scholars find that Chinese intonation is much more complex than the English one. It is because that, on one hand, Chinese is a typical tone language, with each syllable having certain F_0 contour which is lexically given and cannot be changed arbitrarily, while on the other hand, this local F_0 contour must be modified to meet the requirement of intonation. Consequently, some related issues must be raised, such as, what the Chinese intonation is like? Is there any explicit global form for Chinese intonation? And how do the lexical tones integrate with intonation?

1.2.1 Challenging related to the global form of intonation

Is there explicit global form in Chinese intonation? It is a hot issue argued recently.

According to the observation to natural speech, Tseng (1997) pointed that there is only about 70% sentences with pitch declination.

After a series of profound approaches, Xu (1999, 2001) proposed that there is no explicit global form for the intonation of an utterance, and that the surface F_0 declination is determined by the role of multiple sources, such as the downstep caused by L tone, new topic or focus in the utterance, and these effects are parallel.

Shih (2001) conducted an investigation to a set of sentences, which were designed consisting of syllables all with high level tones, so as to avoid downstep effect from low tone features. The experimental data showed a clear declination effect, the syllables early in the sentence having higher pitch level than the later ones. At the same time, her data also showed a sentence length effect where longer sentences have higher starting F_0 values. Of course, Shih also observed the focus effect, which was with pitch expansion on the focused word and steep declination after focus.

Sun's study (2002) assumed the existence of integration among different intonation levels, but did not suppose an entire intonation form for the utterance. Sun suggested that, the F_0 contour in phrase level is concatenated by that in word level, while the F_0 contour in word level is concatenated by that of syllables. The F_0 contour of syllable is the smallest intonation unit, within which different

intonation components are vertically superposed, but not horizontally connected.

Recently, Wang (2003) also investigated the pitch movement by analyzing 600 sentences, which were designed with four lexical tone sequences respectively. Her findings confirmed that there does exist an underlying F_0 declination besides the downstep effects coming from the low tone feature, new topic and focus. She found that, in the declarative sentences without new topic and specific focus, a deep-level pitch prototype gradually descend from the beginning to the end, and as for the sentences with low feature tones or certain focus, the downstep effects make the sentence declination more evident.

1.2.2 Challenges in the relationship between tone and intonation

To explore the working mechanism for each aspect of prosody in Chinese, many contributions have been made, and various controversies were raised as well. Among these challenges, the most prominent issue is concentrated on the relationship between lexical tone and intonation. The key point is how the local tones are integrated with global intonation? In the other words, how the individual syllable or word is said in any intonation without losing its lexical identity?

In respect to this topic, various theories have been proposed since early in last century. The most powerful one is the so-called "small ripples riding on large waves" theory which was first suggested by Yuan-Ren Chao (1933), holding that the tones are integrated with intonation just like small ripples riding on large waves, the relationship between which is a kind of “algebraic sum.” Unfortunately, however, this theory is still remained as controversial. The focus of arguments is how to view the “algebraic sum” in that theory. For example, Shen (1985) argued that the relationship between tone and intonation seemed difficult to be counted as an “algebraic sum”, and pointed that Chao himself had revised the theory in his later research. Therefore, He proposed to revise this theory and suggested (Shen, 1992) that intonation modify the pitch range of tone, and it is manifested through tone’s pitch contour, specifically, the top-line of the tones are modified by stress, while the bottom-line by rhythmic construction. Indeed, Chao himself did make some revision on the theory in one of his later lectures. His original “algebraic sum” theory includes two types of adding, namely, simultaneous adding (*i.e.*, synchronic superposition) and successive adding (Chao, 1933). After that, Chao did give up the idea of successive adding and regarded it as the effect of mood auxiliary word (Chao, 1959). However, the idea of synchronic superposition was retained and confirmed several times in his later papers or books (*e.g.*, Chao, 1968, 1959/1980). Even in his book of *A Grammar of Spoken Chinese* (1968) still briefly interpreted the “algebraic sum” relationship between tone and intonation. For example, he interpreted that, in the sentence of “wo3 xing4 lu4, ni3 xing4 wang2”， the intonation of “lu4” was entirely raised, yet still keeping its falling contour of the 4th tone while the intonation of “wang2” entirely lowered and its rising contour of the 2nd tone well kept. This interpretation means that Chao still persisted in the “algebraic sum” theory in his later stage.

At the same time, Wu (1992-1993 / 1996a) argued that in the case of various prosodic and attitudinal modifications, the frequency and thresholds of local and global tone contours in spoken Chinese sentences are difficult to analyze by conventional methods, such as those used in top- and bottom-line model. Wu (1996b) further interpreted that the relationship between tone and intonation was not difficult to be counted as an “algebraic sum”, because Chao’s “small ripples” riding on “large waves” was result of the increasing of tonal scale rather than the changing of its contour shape. Therefore, Wu suggested that the key point to understand Chao’s theory was to separate the contour from the register of pitch movement. Based on this knowledge, he further proposed a new method of analysis for Chinese intonation. It can be summarized into two main parts, namely, 2-4 syllabic tone sandhi patterns and the “change-key” rules. Specifically, the global intonation of a sentence can be built by connecting relevant tone sandhi patterns, which as a phrasal

F_0 contour, it is relatively constant (Wu, 1990; 1996). Then comes changing the keys of these phrasal units according to their stress status in the sentence. The source of Wu's idea is the music theory, regarding the tone sandhi pattern in a sentence as the main theme of a musical phrase while regarding an overall intonation as the melody in a musical score. Thus the shifting of the frequency range in local tone contours is regarded as the transposition or change-key of the melody in music.

Wu's suggestions greatly inspired us on the understanding to Chao's theory, and enabled us to go deeper to explore how the local tones are integrated with global intonation. Under this enlightenment, Cao (1998, 2002) also conducted several investigations by examining tonal variations in Mandarin, and tried to find some operable why to separate the pitch contour of a syllable from its pitch register. The experimental results confirmed that Chao's "algebraic sum" theory is basically reliable, but needing further improvement. The present paper will describe the improvement made according to those experimental data.

2. Discussion

It is commonly agreed that the surface F_0 of a syllable in Chinese is the result of complex modification coming from multiple factors. Consequently, to recognize the relationships between different components in intonation, one of the prerequisites is to decompose the surface F_0 contour into possibly operational elements. Unfortunately, however, it is a very difficult task.

Based on our basic knowledge, tonal feature of a syllable in Chinese contains two aspects: 1) The time-varying pitch contour of the syllable whose variation is mainly constrained by phonological rules, including lexical tone contrast and the tone sandhi rules; 2) The inherent pitch register of the syllable, for example, in Mandarin, the 1st tone and 4th tone categories are inherent with the H feature of register, while the 3rd tone category with the L register feature. Bearing this knowledge in mind, and under the enlightenment of Chao's and Wu's theories, we managed to recognize intonation structure by decomposing the surface F_0 contour into operational elements.

2.1 The basic skeleton of Chinese intonation in universal aspect

In order to go deeper into the structure of Chinese intonation, Cao (1998) proposed a method to decompose the surface F_0 contour into pitch contour and pitch register, and then applied this method to an investigation. The test materials were recorded from news announcements. Generally, in Chinese, the pitch contour of each syllable or lexical item is observable from its spectrogram, but the pitch register is not explicit. To make the later one observable, we defined the register as the average pitch value that was calculated according to the high point and low point of each syllable or lexical item. At the same time, this method also made possible to define the trend of pitch movement for each phrase or sentence. Practically, it was likened roughly by a line, which was drawing through the register values of the first and last syllables in a certain phrase or sentence.

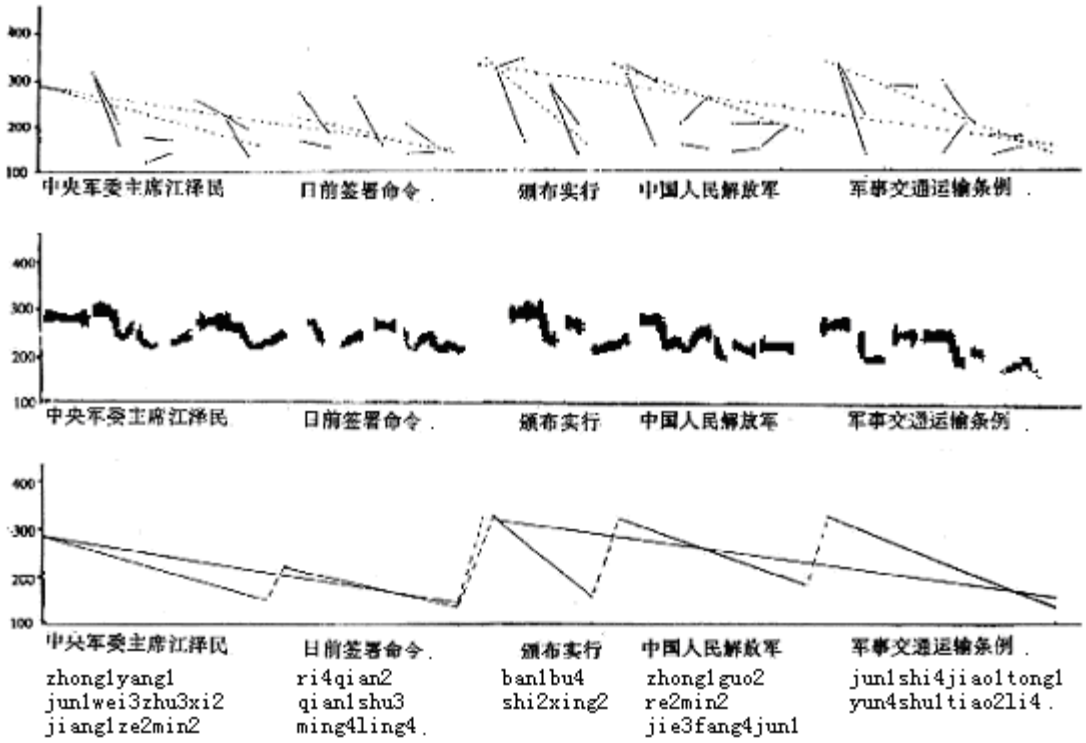

Figure 1: A diagram of the global pitch declination and local pitch variation in the utterance.

Figure 1 gives an example of the diagram in that study. In the Figure, the curves in mid line are the original trajectories of pitch movement of the sample utterances; from which, the lexical tone pattern of each syllable or word can be clearly observed. At the same time, the high point and low point of pitch for each syllable was extracted, then the high point or low point of each word were connected respectively by using short solid lines as shown in the top of the Figure, with those dotted lines roughly representing the trend of register movement for each phrase and sentence. To make it more observable, we connected these dotted lines and extracted them out, then a "large wave" appeared immediately, just like the solid lines shown at the bottom of the Figure. This large wave vividly depicts the undulation of intonation contour. So far, a basic intonation skeleton has been decomposed from the surface F_0 contour, and the pitch declination and its resetting can be observed clearly both in phrase and sentence layer. The trend of register movement obtained here is quite similar to that presented in intonation languages like English. It indicates that there is likely a global form of intonation in Chinese.

To exam whether the situation described above does commonly exist in natural speech, a further study was conducted (Cao, 2002), in which the test materials were selected from a discourse corpus of utterances read by multi-speakers, here four speakers' utterances were examined. The distribution of pitch register of syllables in terms of average value in the utterance is listed in Table 1.

Table 1: Average pitch value (Hz) of the syllable with different tone categories in different positions of speech.

Condition Position	Mean value	1st tone	2nd tone	3rd tone	4th tone	Neutral tone
Sent. start	*196.2*	224.4	182.3	154.8	223.1	--
Sent. mid	*164.7*	182.8	158.1	149.1	171.2	162.3
Sent. end	*113.6*	107.0	115.7	--	118.1	--
Phrase start	*184.4*	190.2	182.6	159.1	205.8	--
Phrase mid	*164.7*	182.8	158.1	149.1	171.2	162.3
Phrase end	*143.7*	143.4	152.0	--	154.9	124.6
Gen. mean	*166.7*	*180.1*	*161.7*	*152.5*	*169.2*	*154.0*

Note: the figures in the lines of sentence mid and phrase mid are the same, all representing the pitch value of the syllable in non-terminal positions in the utterance.

According to the data listed in Table 1, some situations can be observed, which are similar to those described above.

First, in all the cases, the register at sentence start is higher than that of sentence mid and end, and the same is with the phrase level. It means that there does exist declination of pitch register in naturally read discourses, and it is systematically identified no matter in general situation or in the case of different tone categories. The phenomenon observed here indicates how robust the declination mechanism is in real speech, and it is likely the exact basis of intonation skeleton.

Second, such declination is organized as a hierarchy: On the one hand, the declination profile is formed in sentence and phrase level respectively, which is relatively independent in each level, thus to form as different layers. On the other hand, the declination profile of phrases is embedded in that of sentences, thus to form as an organic unity. In fact, in real speech, the size of prosodic phrases is usually different and the smaller ones are embedded in the larger ones. Consequently, the declination profile in real utterance usually has more complex hierarchy.

In addition, as can be seen from the bottom line in Figure 1, these declined trends in different layers are resetted regularly in corresponding prosodic boundaries, thereby leading to the undulation of intonation “large wave” of the utterance. Compared to the pitch movement within a syllable or a word, the pitch undulation referred to intonation is relatively slow and simple. Moreover, The specific declined slope in speech will be further modified by speech mood, focus distribution and rhythmic structure. These modifications are also achieved through register adjustment, whose details will be specified in section 2.3.

Accordingly, as the basic skeleton, Chinese intonation does have a global form in pitch movement. Such underlying skeleton just shows the universal aspect of Chinese intonation, though it is not directly visible as that in intonation languages.

Actually, due to the physiological constraints during speech production, a gradual declination trend of F_0 movement and its resetting must be taken place naturally. Since people usually speak during the expiratory phase in respiratory cycle, the gradual energy losing must lead to the dropping in speaker’s sub-glottal pressure (Liberman, 1967), so it is not surprised that such dropping will result in the F_0 declination.

2.2 The local characteristics of Chinese intonation in special aspect

As for the special aspect, Chinese intonation has more complex local constituents which form as its body and spirit.

As what was mentioned before, each syllable in Chinese has fixed time-varying pitch pattern, which are lexically given and cannot be changed arbitrarily. However, in real context, the exact pitch movement of individual syllables is far deviated from their citation form, whose variation can be summarized into two main levels.

One aspect of the variations is taken place in lexical level due to the phonological constraint, namely, tone sandhi mechanism, which occurrs when syllables form as a word or word-combination. In most cases, the tone sandhi group is disyllabic or trisyllabic chunk. From the mid line of Figure 1, such lexical tone patterns can be seen everywhere over the utterance, with their pitch contours varying as determined by tone sandhi rules (Wu, 1990), thus forming the so-called "small ripples." In addition, within this tone sandhi domain, some inter-syllabic adjustment on pitch register also takes place. Preliminary synthesis test shows that, the specific pattern on such register assignment seems mainly determined by word stress distribution in the context.

Another variation occurs in phrase and sentence level. Actually, that is a further adjustment to the lexical tone pattern, as a result of the modification coming from context, including the effects of stress / accent, rhythm and intonation. Such sort of variations will be discussed later in 2.3.1 to 2.3.6.

2.3 How the special aspect integrates with universal aspect in context?

Tonal feature is the pivot of prosody in a language. It is typically true in tone languages like Chinese. In Chinese, tonal features not only form the basis of intonation and the fundamental essence of lexical tones, but also serve as the decisive factor for stress and accent contrast as well as the indispensable objective property of rhythmic structure. These functions are relatively independent on the one hand, but are closely related to each other on the other. The integrative mechanism may be described as follows.

2.3.1 Relationship between tone and intonation

As mentioned above, in Chinese, both tone and intonation are manifested through the F_0 pattern of individual syllables. The working mechanism can be briefly illustrated with examples shown in Figure 2.

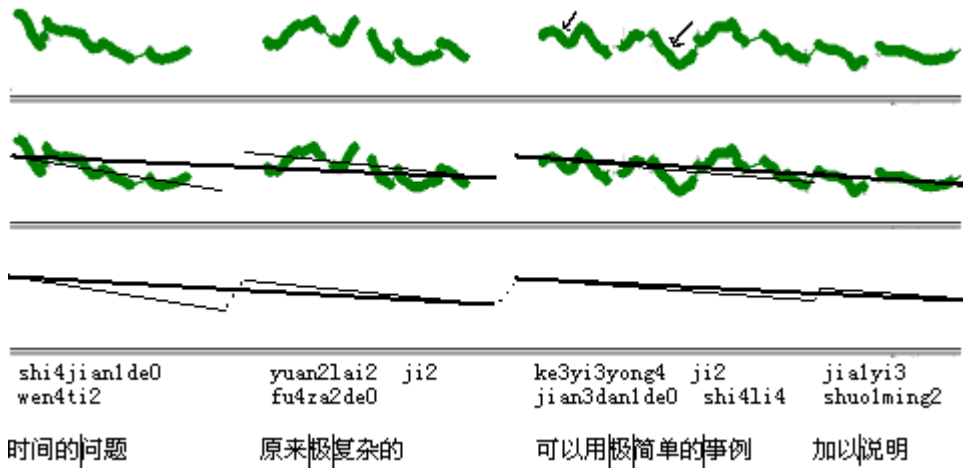

Figure 2: The pitch contour of sentence "世间的问题，原来极复杂的，可以用极简单的事例加以说明".

First, the top line in Figure 2 shows the original pitch contour of the sentence "shi4jian1de0

wen4ti2, yuan2lai2 ji2 fu4za2de0, ke3yi3yong4 ji2 jian3dan1de0 shi4li4 jia1yi3 shuo1ming2"(*"The problem, which is originally extremely complex in the world, can be interpreted through quite simple instances"*). This sentence consists of four prosodic phrases, including eleven lexical tone groups as shown by those characters separated by vertical straight lines, which are here defined as prosodic words.

Then, both the solid lines added on the original pitch contour in the mid and the broken lines at the bottom of the Figure were drawn using the same method described in section 2.1, which represent the trend of pitch movement of the phrases and sentences, and are likened to the intonation "large waves" of the utterance.

From this Figure, the relationship between tone and intonation can be detected.

(1) Executing simultaneously in different ways

From Figure 2, we can see that the basic F_0 pattern for local tones is kept, but is to be aligned following the declined trend of intonation by elevating or lowering their register. For example, in the first phrase, syllables "shi4" and "wen4" belong to the same lexical tone category, which have almost the same falling contour pattern, but different pitch register: the former one is obviously higher than the later one. Similar in the case in the syllables "yuan2" and "za2" in the second phrase, and the two "yi3" in the second sentence, *etc.*, each pair of which have similar pitch contour, but different their register, with the former ones' all higher than that of the latter ones. These phenomena signify that the information of tone and intonation are carried synchronically by the same entity, but executed in different ways, *i.e.*, retaining local lexical distinction by keeping their basic contour patterns, while carrying the information of global intonation through the variation of pitch register.

(2) Superposing each other as an algebraic sum

According to Figure 2, we can observe the specific way of superposition between tone and intonation.

As we mentioned before, in Mandarin Chinese, each lexical tone category has their inherent pitch register. For example, the 1st tone or 4th tone is inherent with H feature, while the 3rd tone with L feature. If taking a closer observation to Figure 2, we find that if a syllable is inherent with a H pitch register, when it occurs at or nearby the peak of "large waves", such as the "shi4", occurring at the peak of "large waves" here, then its register will be further elevated, as a result of the sum of H tone register plus H intonation register. If such kind of H register syllable, like the "wen4", occurs nearby the valley of "large waves", then its register is relatively lowered, as a result of the sum of H tone register minus L intonation register. On the contrary, if a lexical unit is characterized by L register feature, such as the first "yi3" in this sentence, when it occurs nearby the peak of "large waves", its pitch register is relatively elevated, because it is result of the sum of L tone register plus H intonation register; while when it occurs nearby the valley of "large waves", like the second "yi3" here, its register is further lowered, because in this position, both the register of tone and intonation is L, thus the sum must be further reduced.

The situation observed here well agrees with Chao's "algebraic sum" opinion of "positive plus positive is more positive" and "positive plus negative will depend on which one's absolute value is larger." Accordingly, the "algebraic sum" relation between tone and intonation not only exists at the end of phrase or sentence, but also acts over the whole utterance. The key point is that such relation is mainly related to the superposition of their pitch register, instead of that of pitch contour. Of course, the pitch contour of syllables is more or less changed, sharper or flatter depending on the context, but the basic time-varying profile remains relatively unchanged.

2.3.2 Relationship between tone and accent

Tonal variation related to stress/accent is usually presented as pitch accent (prominent) of relevant syllable or word, including pitch range expansion and pitch register elevating. This is a universality occurring in all languages, and Chinese is no exception. However, according to the information obtained from natural speech, tonal variation related to accent in Chinese has its own characteristics.

First, the lexically given pitch pattern of the accented syllable must be retained and even further become more typical than that in corresponding non-accented cases; on the other hand, its pitch register is outstood and pitch range expanded. Therefore, both lexical distinction and accentual information are carried out simultaneously. That is why it was likened to be an “elastic band” effect (Chao, 1922, 1935).

Moreover, there is another feature, which may be the most prominent one, that the specific behavior of pitch accent is closely conditioned by the category of lexical tones. For example, if a syllable is to be accented, then its register feature will be the high the higher, the low the lower. Specifically, if the accented syllable has a tone characterized by H feature, then its pitch register will be significantly higher than that of the syllables around; while if it is lexically characterized by L register, such as the 3rd tone in Mandarin, then its pitch register will be typically lower rather than higher than that of the syllables around. At the same time, accompanied with the register prominent, the pitch range between high point and low point in the syllable is usually extended. For example, in Figure 2, the 3rd tone syllable “jian3” is accented, if compared to another 3rd tone syllable “yi3” which is unaccented. Comparing the parts that pointed by arrows in the Figure, we can observed clearly that the pitch register of “jian3” is typically lower than that of “yi3”, and its pitch range is significantly expanded than that of “yi3”. This feature has been applied to and tested in a TTS synthesis system, whose importance and contribution both to the naturalness and intelligibility has been revealed (Cao & Chen, 2003).

In fact, the pitch accent described above is a kind of prosodic strengthening. This phenomenon is the enlargement of feature contrast both in horizontal composed relationship and vertical stereotactic relationship. If compared with the preceding or following syllables in certain prosodic unit, it is a strengthening of contrast in horizontal aspect. If compared with the corresponding syllables that with the same phonological constituents but in different context, then it is a strengthening of contrast in vertical aspect. The more the syllable is accented, the typical the tonal feature is.

2.3.3 Relationship between tone and rhythm

According to another relevant investigation (Cao, 1999), the lexical tone pattern is one of the bases of rhythmic structure. The rhythmic hierarchy must be manifested through the variation of lexical tones.

First of all, in most of the cases, the basic rhythmic unit in Chinese is 2-3 syllabic chunks, *i.e.*, prosodic words, whose tone sandhi patterns are one of the most important coherent features of the basic rhythmic units.

Secondly, as the demarcate markers between rhythmic units in higher levels, *i.e.*, between prosodic phrases or intonation phrases, pitch resetting is always manifested through lexical tones. Specifically, when a syllable or word is located at the pre-boundary position, *i.e.*, a domain-final syllable or word, its tone pattern is slightly reduced to relatively lower pitch register and smaller pitch range. On the contrary, when the syllable or word is located at post-boundary position, *i.e.*, a domain-initial syllable or word, its tone pattern is always strengthened to the highest pitch register

and largest pitch range, as well as a most typical contour shape. Thus, a pitch resetting is appeared.

Thirdly, rhythmic hierarchy is partly signaled by the boundary tone contrast. Generally, a prosodic phrase final within a sentence is marked by a NON-L boundary tone, while a declarative sentence end is always marked by an L boundary tone pattern. In the NON-L cases, the boundary tone usually has relatively lower but not the lowest register; while in the L cases, its register is the lowest, and the pitch range is absolutely compressed. In terms of this contrast, semantically the meaning of continuity or finishing will be distinguished. Of course, a H feature boundary tone is also found at the end of interrogative sentences.

2.3.4 Relationship between accent and intonation

The relationship between accent and intonation may be explored from the following steps: First, to see how the accent position influences the slope of intonation contour; and then, how the accent and intonation are integrated.

At first, let's look at how the accent distribution influences the slope of intonation contour. To do this, we examined the pitch movement of the sentence "zhe4shuang1xie2 bu4jie1shi0" ("*this pair of shoes is not durable*"), which was uttered by 9 speakers, and with different mood and different focus position. The measured pitch data indicate that the slope of intonation contour does strongly depend on the difference in focus position. To make it visible, a diagram is given in Figure3, where the dashed lines briefly represent the intonation contour in each case, the solid lines represent local register movement, and the dotted lines describe the average pitch level of the sentence in each case.

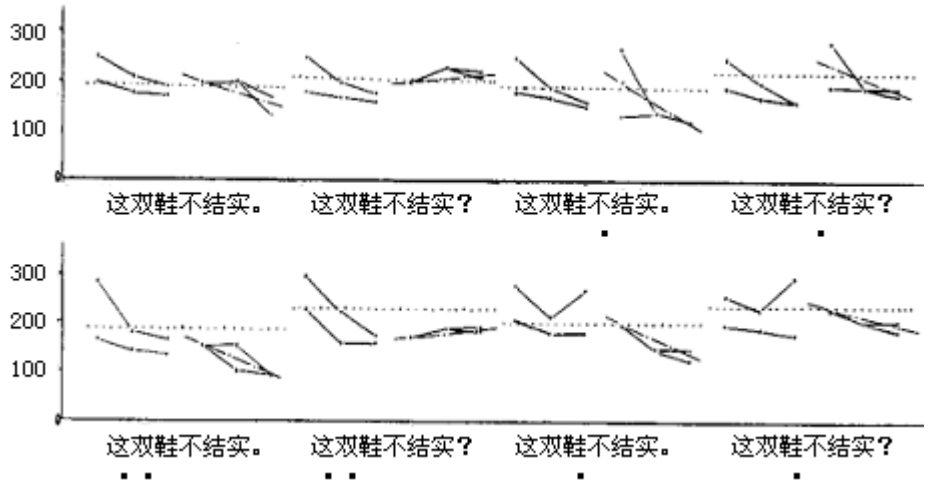

Figure 3: Pitch movement of the sentence "这双鞋不结实 (zhe4shuang1xie2 bu4jie1shi0)" read with different moods and focus positions.

From the picture, first we find a phenomenon that, besides the contour distinction between falling and relative rising, the average pitch level of the sentence uttered in interrogative case is always higher than that of the declarative ones. It is a consistent feature existing in all the cases. However, when the focus position is shifted, the slope of intonation contour will be changed. For example, in the declarative sentence (a) and interrogative sentence (b), there is no marked focus there, usually the word "jie1sh0" is read with a regular (*i.e.*, unmarked or grammatical) accent. We can see that in

this case the intonation contour is falling for (a) and rising for (b). However, in the declarative sentence (c) and interrogative sentence (d), a marked focus is assigned on the word “bu4”, then the falling contour of (c) becomes steeper than that in (a), and the contour of (d) is much more changed as close to a falling one. Here the only difference between (a)(b) and (c)(d) is that, in the case of (a) and (b), the accent is located at the end of the phrase “bu4jie1shi0”, with its elevated register restraining the declination trend of the intonation; while in the case of (c) and (d), the accent is located at the beginning of that phrase, with its elevated register intensifying the declination trend of intonation naturally. Consequently, we may suggest that, the more the accent is approximated to the beginning of the phrase or sentence, the steeper the intonation declination is, and vice versa.

Next, let us exam how the accent and intonation are integrated simultaneously. Generally, due to the declined trend of intonation, it is not difficult to understand that the register of the former unit should be successively higher than that of the latter ones. However, in real speech, the situation is not so uniform, but usually uneven. Such kind of situation can be found from the data in the Table 2.

Table 2: Comparing on the pitch values (Hz) between phrase-start syllables with different positions in a sentence in terms of the mean value calculated from four speakers).

Syllables / Pitch value	“tan4”		“quan2”		“zhi4”	
	mean	sd	mean	sd	mean	sd
Register	214.5	66.7	163.0	45.0	191.9	61.1
High point	254.2	81.6	179.0	48.1	239.3	78.9
Low point	174.8	52.9	147.1	45.1	144.3	46.7

Table 2 lists the register data obtained from a sentence of “tan4tao3yi1xia4 quan2li4zhi4yue1de0 zhi4du4wen4ti2”(“*To inquire into the institution on the constraint of the power”*), which were uttered by four different speakers. The sentence consists of three prosodic phrases, *i.e.*, “tan4tao3yi1xia4”, “quan2li4zhi4yue1de0” and “zhi4du4wen4ti2”. The data show the average value of pitch register, with high point and low point of the syllables “tan4”, “quan2” and “zhi4” respectively. These syllables are all located at the beginning of each prosodic phrase, but distribute at the start, mid and end in the sentence respectively.

Perceptually, the syllables listed here have similar rhythmic prominence, except that the “zhi4” is perceived as the most prominent due to the regular accent rule in Chinese (Zheng *et al.*, 2000). However, according to the data listed in the Table, the values of pitch register, high point and low point of these syllables are systematically different, from which we can observe an interesting phenomenon that, the values of syllable “tan4” is obviously higher than that of syllables “quan2” and “zhi4”, while that of “zhi4” is slightly higher rather than lower than that of “quan2”. This phenomenon indicates that it may reflects an “algebraic sum ” relation between accent and intonation, *i.e.*, on the one hand, the pitch value of the“zhi4” should stand out of other syllables including “tan4”and “quan2” in this sentence so as to meet its accented status; on the other hand, however, because its location is nearby the valley of the intonation “large wave”, its pitch value must be lower than that of “tan4.” Thus, a combination is reasonably resulted from the two effects, *i.e.*, the pitch value of “zhi4” is lower than that of “tan4” so as to carry the information of sentence intonation in this case, but higher than that of“quan2” so as to carry the information of accent. Consequently, I would regard this phenomenon as a kind of simultaneous superposition, too.

To exam whether the phenomenon observed here is a universal significance occurring in natural speech, a further statistical study was conducted based on the data obtained from a large speech corpus, which was labeled according to 48 listeners’ perception, with the acoustic parameters were measured as well (Wang *et al.*, 2000). As the main test objects, eight pairs of phrase-initial

syllables and corresponding phrase-final syllables were selected from this corpus, with the principles of (1) each pair of syllables have the same lexical tone category; (2) phrase-final syllables must be perceptually more prominent than those of phrase-initial ones. The results are summarized in Table 3, from which we can see that the pitch value of phrase-initial syllables is obviously higher than that of phrase-final syllables, though the perceived accent of phrase-final syllables is systematically stronger than that of phrase-initial ones. A statistical test shows that this difference is quite significant (P = 0.01 in the case of high point; P = 0.06 in the case of the register). This result agrees with the one observed from those individual samples described above. It further confirms that there does exist the relation of simultaneous superposition between accent and intonation in common speech.

Table 3: Comparing on the pitch values (Hz) between phrase-initial and phrase-final syllables in terms of the mean value calculated from a labeled speech corpus.

Syllable position	Perceived accent degree		High point		Low point		Pitch register	
	mean	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*
Phrase-initial	0.56	0.29	275.4	72.4	190.0	58.8	232.7	58.1
Phrase-final	0.82	0.46	231.4	44.2	130.6	54.9	181.0	43.5

2.3.5 Relationship between rhythm and intonation

According to our experimental results, an organic integration between rhythm and intonation can be observed. It can be summarized as follows. (1) The basic rhythmic unit is also the basic bearing unit of intonation, the rhythmic hierarchy forming the base of intonation structure. (2) The unique pitch declination of intonation pattern also serves as the coherence of rhythmic unit, and the pitch resetting of related declination is one of the most prominent boundary marker between rhythmic units (Cao, 1999). (3) As one of the markers of hierarchical demarcation between rhythmic units, the contrast of NON-L boundary tone pattern at the phrase-final within a sentence *vs.* L boundary tone pattern at sentence end is the result of the modification from global declination trend. For example, in the sentence of “tan4tao3yi1xia4 quan2li4zhi4yue1de0 zhi4du4wen4ti2” (see the example sentence mentioned in 2.3.4), according to relevant pitch data, the “xia4” in phrase “tan4tao3yi1xia4” and the “de0” in phrase “quan2li4zhi4yue1de0” are all with a NON-L boundary tone there, though both of them are located at the valley of intonation “large waves”; while “ti2” in the phrase “zhi4du4wen4ti2” is with a L boundary tone there. It is because of that the phrases of“tan4tao3yi1xia4”,“quan2li4zhi4yue1de0”and “zhi4du4wen4ti2”stand on different rhythmic layers: The former two stands on the layer within the sentence, which also convey some information of the continuation, so usually with a NON-L tone feature; while the later one is actually located at the end of the sentence, thus must be cued by a L tone feature.

2.3.6 Relationship between rhythm and accent

The organic integration between accent and rhythm may be viewed as follows.

On the one hand, the degree of accent in the utterance is a sort of relative contrast, rather than an absolute opposition. However, such contrast is not an unsystematic behavior, but a well-ordered one based on the rhythmic hierarchy, with the accent in higher rhythmic layer stronger than that in lower layers. In this sense, the degree of accent in the utterance is determined by rhythmic structure.

On the other hand, stress and accent are the important coherent features for the formation of certain rhythmic units. Generally, there must be a relatively accented syllable or word in each rhythmic

unit and every rhythmic layer, which acts as the key net to gather and organize the unaccented syllables around as a rhythmic unity. So in this sense, rhythmic structure is determined by accentual distribution. The number of accent degrees is identical to that of rhythmic layers.

3. A relevant note on possible pre-planning in speech production

Countering to the superposition model, Pierrehumbert (1980) contended that superposition model signifying the global intonation pattern should be pre-planned by the speaker, but such pre-planning seems hard to be believed in cognition.

However, there does exist some interesting phenomena calling for deep thought.

First, according to the study on speech production, Kohno, M. & Tsu Shima, (1989) reported that, the number of syllables in a child's babbling and one-word sentences is never beyond 7. Another approach related to cognation memory (Kohno, M. & Tomoko Tanioka, 1990) found that the span of short-term memory is 7 plus or minus 2 syllables. It means that both the span of speech production and the span of speech perception have a well-working chunking, i. e., 7 plus or minus 2 syllables. It may be a circumstantial evidence for possible pre-planning mechanism in speech production.

Moreover, our speech timing data (Cao, 1999) show that there is an intermediate rhythmic layer in Chinese, where the span of unit length is usually 7 plus or minus 2 syllables, and the unit in such size is generally cohered by a unified pitch declination.

Consequently, it seems that a relatively short-term pre-planning is possible to exist in the process of speech production, but maybe it works only within the domain of a phrase or short sentence. Of course, it is only a sort of association in mind requiring further study.

4. Conclusion

According to the preliminary results obtained from our investigations so far, we would suggest that (1) There does contain an underlying form of entire intonation, which presents as the declined trend of pitch register. This trend is slowly downward and upward alternately due to the resetting at different prosodic boundaries. We suppose that such underlying declination is mainly caused by physiological constraint, and strengthened by the downstep effect of L tone, focus and new topic. The specific declined slope in real speech will vary depending on speech mood, focus distribution and rhythmic structure. (2) Our data reveal that Chao's "algebraic sum" theory is basically reliable, but he mainly concentrated on the last syllable of a phrase or sentence, while ignoring other parts of the utterance. Our data support this theory and further clarify that the "algebraic sum" relationship does work over the whole course of the utterance. (3) The function of local tones, accent, rhythm and global intonation are conveyed by the same substance but in different ways. The integration among them is achieved through the variation of pitch register of individual syllables instead of pitch contour of the syllables.

In summary, Chinese intonation does have different layers and they are hierarchically organized as an organic unity, which can be seen in universal perspective and language-specific perspective respectively. The universal aspect is in the use of falling or rising intonation profile to signal statement or questions and so on, which is similar to that in intonation languages. The special aspect is that, the intonation profile is not presented directly by the pitch curve of individual syllables, but hidden behind these complex curves, which is embodied through the register variation of the syllables. Consequently, as a tonal feature of the syllable, the relative invariance of pitch contour and the variability of pitch register is the key point to understand the relationship between different layers of Chinese intonation.

References

Botinis, A., Granstrom, B. & Mobius, B. (2001). Developments and paradigms in intonation research. *Speech Communication*, 33, 263-296.

Cao, Jianfen (1998). Some aspects on Chinese intonation. *Proceedings of the Conference on Phonetics of the Language in China*, Hong Kong, May 28-30.

Cao, Jianfen (1999). Acoustic-phonetic characteristics on the rhythm of Standard Chinese. *Proceedings of 4th National Conference on Modern Phonetics*, Beijing, China.

Cao, Jianfen (2002). The relationship between tone and intonation. *Zhongguo Yuwen (Chinese linguistics)*, 3.

Cao, Jianfen & Chen, Fangxin (2003). Prediction of prosodic structure based on text information and its application in a synthesis system. *Proceedings of NCMMSC7*, Xiamen, China, Nov.23-25.

Chao,Yuanren (1922). Experimental study of Chinese word tones. *Science*, 7:9, 871-882.

Chao,Yuanren (1933). Tone and intonation in Chinese. *BIHP*, 4:3, 121-134.

Chao,Yuanren (1935). Intonation of National Language. *NLW*, 214.

Chao,Yuanren (1968). *A Grammar of spoken Chinese*. Berkeley and Los Angeles: University of California press.

Chao,Yuanren (1980). *The problem of language*. Beijing: Commercial Press. (Originally appeared as the lectures in Taiwan University, 1959).

Fujisaki, H. (1982) "Modeling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation". *The Proceedings of XIIIth International Congress of Linguistics*, Tokyo.

Fujisaki, H. (1988). A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In Fujimura, O. (eds.). *Vocal Physiology: Voice Production Mechanisms and Functions*. New York: Reven, 347-355.

Kohno, M. & Tsu Shima (1989). Rhythmic phenomenon in a child's babbling and one word sentence. Bulletin No. 191, *The Phonetic Society of Japan*.

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

Liberman, A. M., Cooper, F.S., Shankweiler, D.P. & al. (1967). Perception of the speech code. *Psychological Review*, 74, 431-461.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Ph.D. Dissertation. Cambridge, MA: MIT.

Shen, Jiong (1985). Pitch range of tone and intonation in Beijing dialect. In Lin Tao *et al.* (eds.). *Experimental Phonetic Study of Beijing Dialect*, Beijing: University Press.

Shen, Jiong (1992). An intonation model of Chinese. *Yunwen Yanjiu*, No.4.

Shih, Chilin (1997). Declination in Mandarin. *Proceedings of ESCA Workshop on Intonation: Theory, Models and Applications*. Athens, Greece, September 18-20.

Shih, Chilin (2001). Generation and normalization of tonal variations. *Journal of Chinese Linguistics*, monograph series, no. 17.

Sun, Xuejing (2002). *The Determination, Analysis and Synthesis of Fundamental Frequency*. Doctoral Dissertation, Northwestern University, USA.

Tseng, Chiu-Yu (1997). Prosodic grouping: Suprasegmental characteristics of Mandarin connected speech from a speech database. *Proceedings of the 6th ICCL*, June 18-21.

Wang, Anhong (2003). *Research on the pitch downtrend of intonation in Putonghua*. PhD Dissertation. Beijing: Beijing University.

Wang, Bei, & al. (2000). The pitch movement of word stress in Chinese. *Proceedings of ICSLP'2000*, Beijing, Oct. 17-20.

Wu, Zongji (1990). The basic tonal patterns in Chinese intonation. *The Festschrifts for professor*

Wang Li. Beijing: Commercial Press.

Wu, Zongji (1990). Can polysyllabic tone-sandhi patterns be the invariant units of intonation in spoken Standard Chinese? *Proceedings of ICSLP'90*, Kobe, Japan.

Wu, Zongji (1996). A new method of intonation analysis for Standard Chinese: frequency transposition processing of phrasal contours in a sentence. In G. Fant & al. (eds.). *Analysis, Perception and processing of Spoken language*, Elsevier Science B.V. (Originally appeared in the *RPR-IL/ CASS*, 1992-1993).

Wu, Zongji (1996). Yuan-ren Chao's contribution on the tonal study of Chinese. *Journal of Tsinghua University*, Vol. 11, No. 3.

Xu, Yi (1999). Effects of tone and focus on the formation and alignment of F_0 contours. *Journal of Phonetics*, 27, 55-105.

Xu, Yi (2001). Sources of tonal variations in connected speech. *Journal of Chinese Linguistics*, monograph series, No.17, 1-31.

Zheng, P. & al. (2000). The regular accent in Chinese sentences. *Proceedings. of ICSLP'2000*, Beijing, China, Oct. 17-20.