Issues in Chinese prosody: conceptual foundations of a linguistically-motivated text-to-speech system for Mandarin

Richard S. Lavin

School of Information Science

Kyushu Tokai University

9-1-1 Toroku

Kumamoto, Japan

rick@rslavin.net

Abstract

I examine various controversial aspects of Chinese prosody–tone structure, syllable structure, stress, and intonation–and stress the need to view all of these as interacting systems, aspects of a hierarchical prosodic structure. I examine various proposals at these various levels of the hierarchy and suggest which are most appropriate. Specifically, I suggest the adoption of Bao's version of syllable and tone, and Chen's account of stress. As for intonation, it is still not possible to make any definitive claims regarding an optimal model, but I examine work done by Kratochvil, Shih, and Garding et al, and suggest promising directions for future work.

1 Introduction

I propose to consider here a number of troublesome issues in Chinese prosody. This paper is conceived of as a foundation for a text-to-speech system, though I make no attempt to actually develop such a system at this stage. It is more in the nature of an overview of various approaches that have been taken to date, an attempt to extract broad themes from them, and furthermore an attempt to draw up some kind of "checklist" of features or insights that I feel should be incorporated in any linguistically-motivated model of Chinese prosody. I cast my net fairly widely into the literature, both temporally and in terms of personalities, so as not to discard insights of older or less well-known approaches. An important aim is to develop a model that, while being sufficiently explicit that it could in principal be implemented computationally, makes some kind of sense to phonologists.

My main starting point is fairly obvious to many, though controversial in some circles: that we need to incorporate some notion of hierarchical prosodic structure. For the motivation for this decision, I refer readers to work by Anthony Fox dating from the mid-eighties (Fox, 1985, 1986), and summarized in his more recent book (Fox, 2000), which harks back to work done in the 1950's and 60's. Fox looks at prosodic hierarchies proposed by Hockett (1955), Pike (1967), Togeby (1965) and Halliday (1967): it is his claim that, by factoring out differences in terminology and diagramming conventions and by recognizing the difference between primary and secondary hierarchies and between units and features of units, we can say that they were all proposing essentially the same hierarchy. Further, relating known features of Chinese prosody to their domains of operation we can draw up a table as in Table 1.

FEATURE	UNIT
tone	syllable
stress	foot
tone sandhi	foot or intermediate phrase
intonation	intonational phrase

Table 1: Units and features of units

This is not especially novel, nor is it entirely uncontroversial, but it will serve as a starting point.

What I propose to do is go through these levels and features in turn and survey some of the main proposals that have been made. This will of course preclude any deep discussion of individual controversies. But I believe it is appropriate to do this as, by and large, researchers are concerned with either intonation or other lower-level prosodic features, not both: this means that, though there are fine works available on, e.g., tone structure, syllable structure, intonational endings, etc, not enough has been done to draw all this together into a clear picture of Chinese prosodic structure as a whole.

2 The syllable

Before we consider syllable structure as a whole, in particular where tone fits into the syllable, we need to consider the structure of tone itself.

2.1 The structure of tone

2.1.1 Register

Following Yip (1980) and many other scholars, I consider register to be a feature of tone. Register is, incidentally, a classic example of modern linguists rediscovering something that traditional Chinese scholars took for granted. What is controversial with register is how exactly it fits within the total tonal geometry. Figure 1 shows three proposed configurations. Data examined by Chen (2000) and Bao (1999), most of which has been known of since the time of Chao Yuan-Ren (see Chao, 1968), provides a certain amount of evidence regarding constituency that can help us to decide which of these is in fact correct. Let us review the key points of this argument, drawn from various sandhi processes.

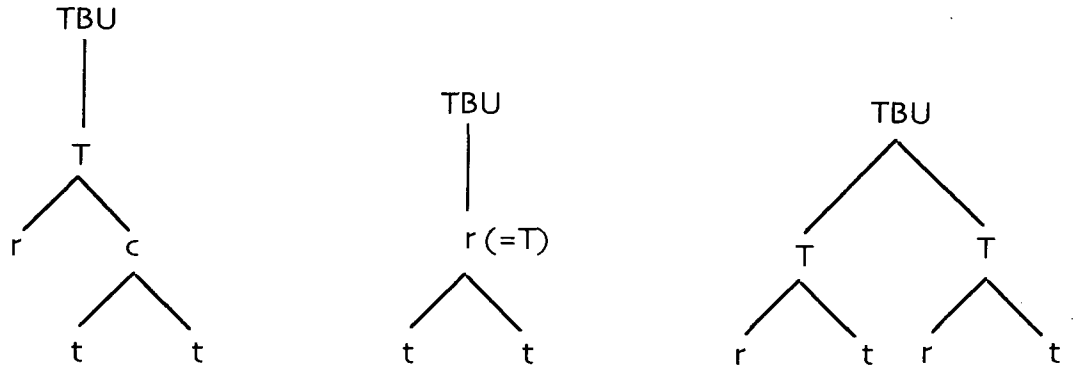

Fig. 1. Tonal geometries proposed by Bao (1999), Yip (1989), and Duanmu (1990), respectively.

TBU = tone-bearing unit T = tone root r = register

2.1.2 Register spread

First, let us consider the Chaozhou word for "warehouse" (lit. "goods storage"), whose first syllable undergoes the following sandhi process:

```
hue  ts'ŋ
l m    l      >     m l    l
```

This and other similar examples lead us to conclude that Chaozhou has a sandhi rule of contour inversion. Consider now the word for "freighter" (lit. "goods ship"):

```
hue  luŋ
l m   h     >     h m    h
```

We could represent the derivation here as:

```
lm  h·   >   ml  h   >   hm  h.
```

(The contour on the first syllable undergoes inversion; then, crucially, the register on the second syllable spreads to the first syllable, displacing its original register.)

2.1.3 Contour shift

In the Zhenhai dialect, the following derivation has been attested (see Rose, 1990, for a fuller account):

```
faŋ   ke
213   441  >  11   334
```

The underlying forms for the tones of these syllables are held to be [l h] and [h l] respectively. Chen (2000) suggests that the derivation proceeds along these lines:

```
l h   h l  >  0   l h  >  l   l h
(0 represents a toneless syllable)
```

(The [h l] contour on the second syllable is deleted through the negative stress effect and the [l h] contour migrates as a whole from the first syllable to the second, leaving the first syllable toneless; finally, a default [l] tone is inserted on the first syllable to avoid leaving it toneless.)

2.1.4 terminal node spread

An example from Mandarin is the behavior of the underlyingly toneless lexical item *de*, meaning roughly "the...one".

```
hong  de                  "the red one"
m h   0    >    l h    m
```

Here, *hong* is analyzed as having H register and [l h] contour.. The toneless *de* takes the default L register. What causes it to be uttered at a medium pitch rather than a low one is the spread of the first syllable's [h] to the second syllable.

2.1.5 whole tone shift

Wenzhou dialect has an idiosyncratic system of disyllabic sandhi in lexical items, whose motivation is rather complex and can only be explained by recourse to the phonological categories of Middle Chinese. The details need not concern us here, but let us consider the word *wenti*, meaning "question".

```
wen ti
 l   ml    >    hm  ml
```

Consider now the multi-syllabic compound for "radio receiver", semi-literally "wireless telephone tube", literally "[[no-wire]-[electric-word]]-tube":

```
wu xian dian hua tong
ml  hm    l   l    ml  >  hm    0    0    0    ml
```

Here, the final [l ml] sequence triggers the disyllabic sandhi mentioned above, despite the grammatical parsing of the compound. The part that interests us is that, finally, this [hm] tone (i.e. a hl contour in H register) migrates across to the first syllable, leaving all intervening tones toneless.

2.1.6 Conclusions for tonal geometry

If indeed whole tones, register only, whole contours, and individual terminal nodes can all variously be involved in phonological processes as constituents, then we can, at least provisionally, adopt Bao's model of tonal geometry.

2.2 Syllable structure

2.2.1 Where does tone fit?

fanqie language games, in which a single syllable is transformed into two different syllables, are rich sources of data when considering what elements of Chinese phonology are actually constituents, since it is assumed that the phonological operations involved can operate only on constituents. More detailed data are provided in Bao (1999) and Lavin (forthcoming), but I present a summary in Table 2:

May-ka	traditional rhyme
Mo- pa	all rhyme segments
Ma- sa	tone;
La-pi	nucleus;
Man-t'a	coda

Table 2: fanqie games and affected constituents

A key point here is that the traditional rhyme (i.e. tone plus all segmental information excepting the onset), and the rhyme as conceived in predominantly segmental approaches (i.e. without the tone) are both constituents. We shall call these the *rime* and *segmental rime*, respectively. This suggests a syllabic template as shown in Figure 2:

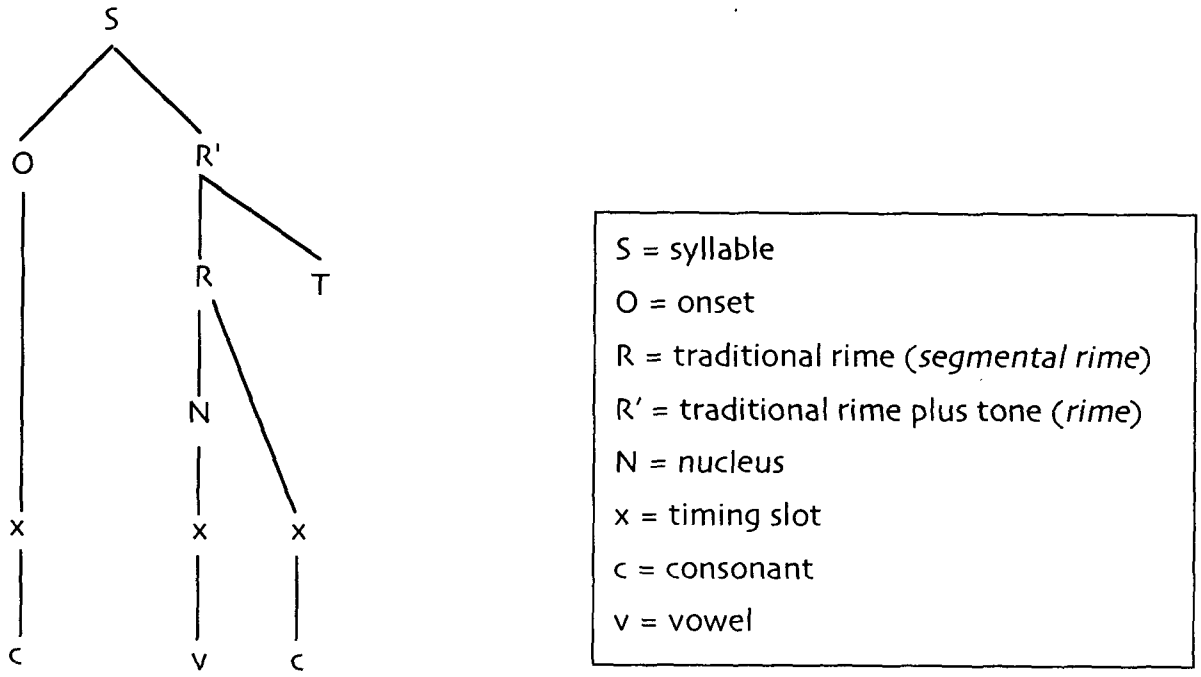

Figure 2: Overall syllable structure

3 Foot structure and stress

Native speaker judgments regarding stress tend to be much less clear-cut for Chinese than for English. This elusiveness is no doubt a primary cause of the fundamental disagreements amongst scholars of the nature of Chinese stress. Table 3 is an abbreviation of my modified version (in Lavin, forthcoming) of Chen's (2000) summary of various accounts.

Chen (2000) offers the most convincing description of Mandarin stress, one that

Uniformly right-prominent	Chao (1968), Yip (1980)
Free stress: predominantly iambic; some lexically marked trochees	Li (1981), Yin (1982), Hoa (1983)
Phrases mostly iambic; root compounds indeterminate	Kratochvil (1964)
Basically trochaic; trochaic reversal causes some iambs	Chang (1992)
Non-head stress	Duanmu (2000, etc.)
Stressability hierarchy: High > Low; Falling > Rising	Meredith (1990)
Indifferent: no lexical stress	Gao & Shih (1963), Duanmu (1993)

Table 3: theories of Chinese stress

simultaneously explains the variability in observed stress patterns and also the pattern of application of tone 2 sandhi, which in certain contexts "changes tone 2 to tone 1 after tone 1": Mandarin is generally right-prominent, but is susceptible to iambic reversal to avoid stress clash; tone 2 sandhi applies only to metrically weak positions, and only within feet (*minimal rhythmic unit* in Chen's terminology). The following examples illustrate these principles:

```
bo (1) luo(2)     "pineapple"
 H      MH
```

Here, owing to general right prominence, *luo* is metrically strong, so tone 2 is realized fully. However:

```
[bo   luo]   kuair(4)    "pineapple cubes"
 H     HH      HL
```

Here, Tone 2 has been "changed to Tone 1", i.e. M > H / H _ H), because stress has shifted from *luo* to *kuair*, leaving *luo* weak and susceptible to sandhi.

```
[Lao(3)  Jin(1)] [chan   jiu]    "Old Jin craves liquor"
   L       H        MH    LH
```

Here, the third syllable *chan* does not undergo Tone 2 sandhi, despite being metrically weak, because it belongs to a different foot than the potential trigger, *Jin*.

Kratochvil (1998) describes iambic rhythm as being inherent, and trochaic rhythm as being largely a consequence of the second syllable of many compounds being suffixes or other relatively contentless syllables, which are particularly liable to negative prominence effects.

4 Intonation

While there is a long history of study of lower level prosodic features such as tone and tone sandhi, intonation has traditionally been comparatively neglected. Some recent works, however, show advances in our understanding of intonation.

4.1 Features

4.1.1 Channelling

This is the term that Kratochvil (1998) uses to describe the "overall guidance...in a breath group...to provide the limits for the execution of the primary and secondary prosodic modifications". He hints that there is a one-to-one correspondence between types of channel and (syntactic?) type of sentence. Obviously, it is not difficult to interpret the basic grid lines of the Lund model as being an explicit implementation of this observation.

4.1.2 Focusing

This is defined by Kratochvil as "a momentary enlargement of a channel...for signaling such features as contrastive stress". This is also handled in the speech acts part of the Lund model: "Part in focus expanded" (Garding et al, 1983). Garding et al also implement a compression of the pitch range after the focused part. Shih (2000) backs this up, but mentions differences between speakers. He also points out the effects of lexical tone: if a low tone follows the focus, then the compression is immediately after the focus; if high tone sequences follow the focus, the effect is gradual.

4.1.3 Declination

This could be subsumed into channeling. Shih (2000) gives the best account of declination, showing that declination sets in near the beginning of a sentence, and that it tends to be faster near the beginning than at the end. He also shows that people tend to make a rough calculation of the length of a sentence before production, and that longer sentences will tend to start higher than shorter ones; however, the rate of decline does not seem to vary significantly. These observations allow him to express declination in terms of three parameters: initial value, rate of decline, and asymptote value.

4.2 Broad typology

To state the obvious, the fact that all models of intonation have to account for is that all the various factors mentioned above–tone, tone sandhi, stress, etc.–come together in some way to produce a measurable F0 contour that unfolds in real time. In other words, in some sense the total contour "consists of" individual tones, stress patterns, intonation, etc. It is the meaning of this term "consists of" that is at issue.

It is fairly uncontroversial that a degree of abstraction is in order to deal with the fact that some segments are intrinsically lower than others and that unvoiced segments cause breaks in the F0 contour. It is trivial conceptually (though not computationally) to factor out these "disruptions" in our phonologically ideal contours. Beyond that, however, almost everything is controversial.

For a broad perspective on the specification of prosody, we revisit Fox (2000). Fox characterizes the SPE method of assigning stress patterns in English, McCawley's (1968) method of deriving the pitch patterns of Japanese, the Liberman and Prince (1977) way of building metrical trees, and Lexical Phonology (Kiparsky, 1982) all as being examples of *bottom-up* approaches.

The main weakness of this class of approaches, according to Fox, is that, though it is partially true that the whole is determined by its parts, it is also true that "features of larger units provide the setting for those of smaller units", and such an approach cannot incorporate this insight.

The opposite class of approach is, of course, the *top-down* type, of which the Lund School approach is a prime example. In this model, sentence and major phrase boundaries first determine the overall tonal grid, and lower-level prosodic features such as stress, pitch accent, and tone are superimposed on the overall sentence intonation. A drawback is that, though the model can generate realistic contours, and is therefore a valid practical synthesis system, it does not specify what elements of the intonation contour are phonologically relevant and is therefore of limited applicability as a phonological model.

Another aspect of prosody, mentioned above, is that utterances take place in real time; therefore a *left-to-right* model aptly characterizes the "basically linear nature of the speech-event itself". Incorporating this left-to-right principle does not of course exclude the top-down or bottom-up principles. Indeed, it is a common feature of autosegmental phonology that tones

and TBUs are associated left-to-right, and also assumed in most accounts of downstep and declination that these are essentially left-to-right processes.

It is undoubtedly true that utterance production is, to an extent, an on-going, improvisatory process. It also appears to be true that speakers incorporate their knowledge of approximate expected length of an utterance into even the early stages of its production. This results in a tendency for the end point of all utterances, irrespective of their length, to be at approximately the same height ('t Hart, Collier, and Cohen, 1990:134).

4.3 Linear/Superpositional; Tone Sequence/Contour Interaction

Ladd (1983) suggests that intonation models can be divided into two types: *tone sequence* and *contour interaction* approaches. This distinction is restated in Hirst & Di Cristo (1998): In a tone sequence (TS) approach, "The pitch movements associated with accented syllables are themselves what make up sentence intonation", whereas in a contour interaction (CI) approach "an intonation contour is the result of the superposition of contours defined on different hierarchical levels". Rossi (2000) terms these two approaches linear and superpositional, respectively.

CI frameworks are stereotypically associated with predominantly engineering approaches, while TS theories, in the broad Pierrehumbert tradition, are more "phonological" in character. They offer the tantalizing possibility that phonological primitives (e.g. tonal segments) are the same in all languages, and that they differ only in how (i.e. at what level) they fit in the prosodic hierarchy. Hirst & Di Cristo's (Eds.) collection of descriptions of twenty languages is certainly promising in this regard. However, it is probably significant that two scholars–including Kratochvil treating Mandarin Chinese–were unwilling to use this framework. And, bearing in mind the popular conception (after Chao Yuan-Ren) of syllabic tones as ripples riding on the waves of intonation, TS frameworks do seem somewhat counter-intuitive. In particular, when the inherently rising tone 2 falls as a result of sentence intonation, it is difficult to see any sensible way of explaining this in terms of a Low tone appended to the end of the Tone Group.

5 Conclusion

In the space available, it has been possible to offer only the most cursory examination of a small number of interesting aspects of Chinese prosody, yet it is my hope that this paper has served as a kind of bird's eye view of the field for those engaged in research in one corner or other of the field who periodically may wish to take a broader view.

References

Bao, Z. 1999. *The Structure of Tone*. Oxford University Press, New York.

Chang, L.M.-C. 1992. *A Prosodic Account of Tone, Stress, and Tone Sandhi in Chinese Languages*. Ph.D. University of Hawaii.

Chao, Y.R. 1968. *A Grammar of Spoken Chinese*. University of California Press, Berkeley.

Chen, M.Y. 2000. *Tone Sandhi : Patterns across Chinese Dialects*. Cambridge Studies in Linguistics ; 92. Cambridge University Press, Cambridge, UK ; New York.

Duanmu, S. 1990. *A Formal Study of Syllable, Tone, Stress and Domain in Chinese Languages*. Ph.D. dissertation. Massachusetss Institute of Technology.

Duanmu, S. 2000. *The Phonology of Standard Chinese*. Phonology of the World's Languages. Oxford University Press, Oxford ; New York.

Duanmu, S. 1993. Rime Length, Stress, and Association Domains. *Journal of East Asian Linguistics,* 2:1-44.

Fox, A. 1985. Aspects of Prosodic Typology. *Working Papers in Linguistics & Phonetics (University of Leeds),* 3:60-119.

Fox, A. 1986. Dimensions of Prosodic Structure. *Working Papers in Linguistics & Phonetics (University of Leeds),* 4:78-127.

Fox, A. 2000. *Prosodic Features and Prosodic Structure : The Phonology of Suprasegmentals*. Oxford Linguistics. Oxford University Press, Oxford.

Gao, M., and A. Shi. 1963. *Yuyanxue Kailun [Introduction to Linguistics]*. Zhonghua Shudian, Beijing.

Gårding, E., J.-L. Zhang, and J.-O. Svantesson. 1983. A Generative Model for Tone and Intonation in Standard Chinese. *Working Papers, Linguistics-Phonetics, Lund University,* 25:53-65.

Halliday, M.A.K. 1967. *Intonation and Grammar in British English*. Ed. linguarum Janua. Mouton, The Hague.

Hart, J.T., R. Collier, and A. Cohen. 1990. *A Perceptual Study of Intonation : An Experimental-Phonetic Approach to Speech Melody*. Cambridge Studies in Speech Science and Communication. Cambridge University Press, Cambridge, England ; New York.

Hirst, D., and A. Di Cristo, eds. 1998. *Intonation Systems : A Survey of Twenty Languages*. Cambridge University Press, Cambridge ; New York.

Hirst, D., and A. Di Cristo. 1998. A Survey of Intonation Systems. *Intonation Systems: A Survey of Twenty Languages*. Eds. D. Hirst and A. Di Cristo. Cambridge University Press, Cambridge, 1-44.

Hoa, M. 1983. *L'accentuation En Pekinois*. Editions Langages Croises, Paris.

Hockett, C.F., and I.J.O.A. Linguistics. 1955. *A Manual of Phonology*. Indiana University Publications in Anthropology and Linguistics. Memoir 11. Waverly Press, Baltimore.

Kiparsky, P. 1982. From Cyclic Phonology to Lexical Phonology. *The Structure of Phonological Representations*. Eds. H. van der Hulst and N. Smith. Foris, Dordrecht, 131-75.

Kratochvil, P. 1998. Intonation in Beijing Chinese. *Intonation Systems: A Survey of Twenty Languages*. Eds. D. Hirst and Al. Di Cristo. Cambridge University Press, Cambridge, 417-31.

Kratochvil, P. 1964. Syllabic Stress Patterns in Peking Dialect. *Archiv Orientalni,* 32(3):383-402.

Ladd, D.R. 1983. Peak Features and Overall Slope. *Prosody: Models and Measurements*. Eds. A. Cutler and D. R. Ladd. Springer, Heidelberg.

Lavin, R.S. forthcoming. Aspects of Chinese Prosody: Tone, Stress, and Syllable Structure. *Bulletin of School of Information Science, Kyushu Tokai University,* 2 (2001).

Li, W. 1981. Shilun Qingsheng He Zhongyin [on Neutral Tone and Stress]. *Zhongguo Yuwen*:35-40.

Liberman, M., and A.S. Prince. 1977. On Stress and Linguistic Rhythm. *Linguistic Inquiry,* 8:249-336.

Mccawley, J.D. 1968. *The Phonological Component of a Grammar of Japanese*. Monographs on Linguistic Analysis, No. 2. Mouton, The Hague, Paris.

Meredith, S. 1990. *Issues in the Phonology of Prominence*. Ph.D. dissertation. Massachusetts Institute of Technology.

Pike, K.L. 1967. *Language in Relation to a Unified Theory of the Structure of Human Behavior*. Janua Linguarum. Series Maior 24. 2nd rev. ed. Mouton, The Hague.

Rose, P. 1990. Acoustics and Phonology of Complex Tone Sandhi. *Phonetica,* 47:1-35.

Rossi, M. 2000. Intonation: Past, Present, Future. *Intonation: Analysis, Modelling and Technology*. Ed. A. Botinis. Kluwer Academic, Dordrecht; Boston; London, 13-52.

Shih, C. 2000. A Declination Model of Mandarin Chinese. *Intonation: Analysis, Modelling and Technology*. Ed. Antonis Botinis. Vol. 15. Text, Speech and Language Technology. Kluwer Academic, Dordrecht, 243-68.

Togeby, K. 1965. *Structure Immanente De La Langue Française*. Langue Et Langage. Larousse, Paris.

Yin, Z. 1982. Guanyu Putonghua Shuangyin Changyongci Qingzhongyin De Chubu Kaocha [a Preliminary Study of Stress in Disyllabic Expressions in Putonghua]. *Zhongguo Yuwen*:168-73.

Yip, M.J.W. 1989. Contour Tones. *Phonology,* 6:149-74.

Yip, M.J.W. 1980. *The Tonal Phonology of Chinese*. Indiana University Linguistics Club, Bloomington, Indiana.