

Towards a Pan-Mandarin System for Prosodic Transcription

Shu-hui Peng, Marjorie K.M. Chan, Chiu-yu Tseng,
Tsan Huang, Ok Joo Lee, and Mary Beckman

11.1 Introduction

This chapter describes the initial stages of development of a Pan-Mandarin ToBI system. We will review the salient prosodic characteristics of Mandarin, with particular attention to the range of variability within a common structural core, and then propose an initial codification of conventions for marking prosodic structure in two standard varieties and one regional variety of the language. Before we begin, however, we must at least pose two important prior questions. First, is a Pan-Mandarin ToBI feasible? Second, if feasible, is it desirable?

The first question arises because of the sheer size of the language. Mandarin is spoken as a first or second language by well over a billion people. Furthermore, it has had a long history of being spoken over large geographical areas as a standard language or as a local language. As a result, Mandarin encompasses many different varieties, including three national standards (Guoyu in Taiwan, Putonghua in Mainland China, and Huayu in Singapore) as well as many regional varieties, such as Rugaohua, discussed in section 11.2.5. Furthermore, a large population of Mandarin speakers speaks more than one variety of Mandarin. Many others speak a variety of Mandarin and some non-Mandarin Chinese or other language variety. Code-switching often occurs, at many different levels, including prosody. The enormous geographical range and the interaction with different substrate languages at the edges of its range lead to variability in syntax, in lexicon, and especially in phonology, including the phonology of tone, stress, and prosodic grouping. The short answer to the first question, therefore, is that we cannot know until we try.

At the same time, we think that this variability makes Mandarin an ideal test bed for developing the ToBI framework into a more powerful descriptive tool. While a Pan-Mandarin ToBI system may be too ambitious, this is our goal. The current Mandarin ToBI system is being designed to describe the prosodic structure and inventory of tones and other structure-marking elements in Putonghua, Guoyu, and several regional varieties. Eventually, we hope that it can be extended to cover all varieties. In developing the system, then, we need to specify how to expand it incrementally to accommodate our expanding knowledge of regional varieties as well as of the standard varieties. With the right accommodation, the system should also be able to describe interactions (such as code-switching events) between different varieties of Mandarin and perhaps between Mandarin and other varieties of Chinese (and other languages) in different social contexts. The accommodation to variation and code-switching is of course a more general issue for the ToBI framework (see, e.g., Bruce, this volume; Grice et al., this volume). Thus, a general answer to the second prior question is a resounding affirmative.

We can also answer the second question more specifically in response to the aims of this book. The Mandarin varieties of Chinese are a particularly fruitful source of data for typological comparison across prosodic systems. There are several reasons for this. First, all varieties of Mandarin have lexical tone contrasts, and all show a fairly dense specification of tone (albeit none as dense as in Cantonese — see Wong et al., this volume). At the same time, many Mandarin varieties have metrical structure reminiscent of the stress systems seen in the Germanic languages, with tone specification constrained by syllable prominence at both the lexical and the phrasal levels. Thus, just as Cantonese nicely illustrates the interaction between lexical tone and the “purely intonational” use of tones as pragmatic morphemes that mark phrase edges, Mandarin illustrates the interaction between lexical tone and the “accentual” use of tone as a stress marker.

Mandarin is a fruitful source of data for typological comparison also because there is enough variability just within Mandarin itself to look at the effects of minimal differences in just one aspect of the prosodic system while keeping the lexicon and the rest of the prosodic core intact. For example, there is variability across varieties in the incidence and distribution of unstressed (toneless) syllables that is fully comparable to that seen in the Germanic languages. Just as Swedish differs markedly from English and German in the distribution of “de-accented” words in running speech, Guoyu (the standard Mandarin of Taiwan) differs markedly from Putonghua (the standard of the People’s Republic of China) in the distribution of “neutral tone” syllables (Tai 1978).

Another point of difference across Mandarin varieties lies in the evidence for categorically marked prosodic units above the syllable. For example, all varieties of Mandarin have “tone sandhi” — i.e. phonologically- or lexically-conditioned tonal alternations that are vaguely reminiscent of the phenomena called “tone sandhi” in the Wu varieties of Chinese, such as Shanghainese (Jin 1985), Danyang (Chan 1991) and Wuxi (Chan & Ren 1989). However, unlike in the Wu varieties, the relationship between tone sandhi and prosodic structure in standard varieties of Mandarin is quite controversial. Some earlier accounts emphasised the blocking of tone sandhi at certain syntactic constituent edges (e.g., Cheng 1973), which might suggest a Wu-like phonological phrase smaller than the intonational phrase as the domain of tone sandhi (see, e.g., Liao 1994). This characterisation is probably accurate for tone sandhi in the regional varieties that border on the Wu-speaking regions of China. However, for the standard varieties, other more recent accounts emphasise rhythmic constraints, and prominence alternations within the “foot” (e.g., Shih 1986, 1997; Zhang 1988; Hsiao 1991). In short, there are no unifying commonalities such as those that define the “tone sandhi group” as a shared prosodic unit in all Wu varieties despite the large and

salient differences in details of tone shape and initial- versus final-syllable dominance. We might say that the phonological characterisation of delimitative markers for any shared core of higher-level prosodic groupings across varieties is complicated by the relationship in the standard varieties between tone, on the one hand, and stress or “foot-level” rhythmic grouping, on the other.

In this chapter, then, we will briefly describe the salient characteristics of the Mandarin varieties of Chinese that a ToBI annotation system should capture. We will give due emphasis to comparison across different varieties of Mandarin, in addition to the obvious points of reference to other non-Mandarin varieties of Chinese and to other languages that have been described within the ToBI framework. We will also outline how the ToBI framework has been adapted at two different sites to develop annotation conventions for tagging spoken language databases. Since the research focus differs somewhat between the two sites, with one site working to build the requisite infrastructure for speech synthesis and recognition, and the other site developing corpora to explore sociolinguistic variation in prosody across different varieties and styles, Mandarin ToBI also provides a useful illustration of how basic research in prosodic typology can inform other areas of linguistics and speech science.

11.2 Salient characteristics of Mandarin

11.2.1 The scope of Mandarin

As noted above, one of the characteristics of Mandarin that is especially challenging for the ToBI framework is its size. Some 70 percent of the Chinese-speaking population of the world are speakers of Mandarin. There are more than 800 million native speakers of Mandarin just in the People’s Republic of China, and at least half that many speakers of Mandarin as a second language (see Grimes 1996). Mandarin is spoken in all regions of China north of the Yangtze River and in

large regions south of it in southwestern China. This geographical distribution is three-quarters of the entire primarily Chinese speaking territory of the PRC (Yuan 1960, 1989). Thus, the Mandarin-speaking region of mainland China is large enough to encompass considerable variability. Yuan (1960, 1989) classifies Mandarin into four main varieties based in part on geography and in part on phonological differences: Northern Mandarin, Northwestern Mandarin, Southwestern Mandarin, and Jianghuai Mandarin. Li (1985, 1989a, b) designates the Chinese spoken in Shanxi Province in northern China as non-Mandarin, and then subdivides the rest of Yuan's Mandarin into 7 major regional varieties based on modern reflexes of the so-called 'entering tone' (*rusheng*). Neither of these classifications includes the variability that has emerged among speakers of Mandarin in other parts of the world, such as Taiwan, Singapore, Indonesia, Thailand and other parts of Southeast Asia, the United Kingdom, North America, and South Africa.

Across the Mandarin-speaking world, Mandarin is in contact with many other varieties of Chinese, as well as of non-Chinese languages. This contact increases the variability even more. For example, the Jianghuai varieties share many phonological features with the neighboring Wu varieties of Chinese. They have even been described as being underlyingly Wu with a Mandarin superstratum (Ting 1966).

Even discounting such variability across regional varieties, Mandarin is still a very heterogeneous language. Whether speaking Putonghua as a first or a second language, speakers of this standard variety will show regional features and differences associated with age and social class. Even the prescriptive Putonghua of trained broadcasters differs from the "original" source of the standard, Beijinghua (Hu 1987). And Guoyu, the standard Mandarin of Taiwan, shows features suggesting a Min substratum. This is not surprising, since Taiwanese, the native language of somewhere between 70% and 80% of Taiwan Chinese, is a Min variety (see Cheng 1985, Feifel

1994). Cheng (1985) also observes that, among the especially influential and economically powerful non-Taiwanese speakers of Mandarin in Taiwan there are many Wu speakers, including the former ruling family. Thus, as a result of contact with Taiwanese, Wu, non-Beijing varieties of Mandarin, and so forth, these two standards today – Putonghua and Guoyu – are quite different. Tai (1976, 1977, 1978), for example, systematically documented phonological, syntactic, lexical, and stylistic differences between Putonghua and Guoyu that already reflected significant changes after only the first quarter-century of political separation. Thus, even if we limited Mandarin ToBI to dealing only with “standard” Mandarin, there would be much variability to cover.

11.2.2 Lexical tone

Lexical tone is another salient characteristic of modern Mandarin. As in other varieties of Chinese, every syllable in a Mandarin utterance can be identified as a free or a bound morpheme. And almost every syllable — even one that is a bound morpheme — is lexically specified for one of several phonemically contrasting tones. Tone specification is so prevalent in the lexicon that the exceptional syllables are traditionally described as having a tonal specification — for “neutral” tone (see section 11.2.3).

The number of tones in contrast is often four, as in the standard varieties, although it can be as few as three, as in the Yantai and Qingdao varieties spoken in Shandong Province (see Qian 1982, for Yantai, and Qingdaoshi Shizhi Bangongshi 1997, for Qingdao). Although there are always exceptions, due to “dialect borrowing” at various stages in the history of the language, it is often possible to identify a fairly simple set of correspondences between cognate forms. However, even when two varieties show such simple correspondences, they can differ markedly in the phonetic tone shapes for cognate morphemes in citation form. Moreover, even between the

standard varieties, we see divergence in phonetic shape. Because of this variability, contemporary linguists often refer to the lexical tones as “Tone 1”, “Tone 2”, “Tone 3” and “Tone 4” rather than with more descriptive terms such as “high tone”, “rising tone”, and so on, and we will often adopt this terminology in talking about the standard varieties, where the cognate sets are quite well-behaved. (Old Mandarin also had four tones, and classically educated linguists sometimes use traditional, philological terms, such as *yin-ping* versus *yang-ping* for Putonghua Tone 1 and Tone 2. However, the correspondence between the four tones of Old Mandarin and those of the modern standard varieties is not simple — cf. Stimson 1966, for example, for a discussion of the modern reflexes in Beijing Mandarin. Hence, we will resort to these philological terms only for describing the tones of “checked” *rusheng* syllables in regional varieties that retain vestiges of the syllable-final stops of Middle Chinese.)

The four tones in standard Mandarin are described variously by different linguists. Some of these differences are due merely to the phonological framework assumed. Others reflect real contextually conditioned variation. Chao (1948, 1968) lists the citation tone shapes as “high-level” for Tone 1, “high-rising” for Tone 2, “low-dipping” for Tone 3, and “high-falling” for Tone 4. He further notes that in non-citation context, the third tone surfaces as “high-rising” before another third tone in tone sandhi contexts, and as “low-falling” elsewhere. Kratochvil (1968), by contrast, lists the four tones as “high”, “rising”, “low”, and “falling” — a phonemicisation that is amenable to a two-level Autosegmental analysis as H, LH, L, and HL. Only in the phonetic description does he identify the third tone as having a dipping pitch contour in citation forms.

Table 11.1 shows a traditional alphabetic phonetic representation in terms of Chao’s (1930/1980) tone numbers (with five levels), taking Chao’s “elsewhere” shape as the basic form for Tone 3, as in Shih (1986, 1988). Note that Shih’s decision regarding the “basic” shape for Tone 3

was intended specifically for Guoyu. This is in keeping with Tai's (1978) observation that even the citation form of Tone 3 tends to be produced as [21] in Guoyu, rather than with the full dipping tone that Chao chose as the "basic" citation form shape. Both Tai and Shih, then, assume that the Guoyu tone system differs from the Putonghua system primarily in the distribution of the Tone 3 allotones. (However, see Hartman (1944) who describes the dipping variant of Tone 3 only for "loud-stressed syllables occurring finally", and he was writing before the emergence of Guoyu as an independent standard.)

More recent work has uncovered other differences, which may have emerged only in the quarter century since Tai's (1978) treatise. A corpus of productions by a representative younger speaker described in Fon and Chiang (1999) suggests two differences. First, whereas in Putonghua, a Tone 3 syllable in its "full" [214] form is the longest of all citation form types, the corresponding "dipping" variant of Tone 3 in Guoyu is shorter than Tone 2. Thus, the relationship between the falling and the dipping variant in Guoyu cannot be ascribed to tone truncation. Also, Tone 2, which in Putonghua is typically a rising tone in citation form, in Taiwan Mandarin is another "dipping" tone, which might be transcribed as [323]. (The corpus shows timing differences that make the two dipping tones more different from each other than these alphabetic transcriptions suggest, and these timing differences are perceptually salient, as shown by Fon 2000.) Thus, in the half century since Mandarin was enforced as the standard language of Taiwan, Guoyu has differentiated itself from Putonghua in ways that may eventually be as drastic as the differences between Putonghua and regional varieties of Mandarin within the People's Republic of China.

For the sake of comparison, Table 11.2 gives a traditional alphabetic representation for the Rugaohua citation-form tones. (Rugaohua is the regional variety of Rugao and neighboring counties in Jiangsu Province, and classified as a subgroup of Mandarin, Jianghuai Mandarin, with

about 67 million speakers — see Li 1985, 1989a, b.) There are two “extra” lines in the table, illustrating the two *rusheng* tones on “checked” syllables. As in all varieties of Jianghuai Mandarin, Rugaohua still retains some vestige of the syllable-final stops of Middle Chinese, although the three-way place contrast of Middle Chinese has been lost, with historic /p/, /t/, /k/ all replaced by final glottal stop, as in the Wu varieties of Chinese. Disregarding the correspondences for these checked syllables, and comparing only the four types in syllables with all-sonorant rhymes, we can see that Tone 2 and Tone 3 are similar to the corresponding tones in Putonghua, but Tone 1 and Tone 4 are not.

Insert Table 11.1 about here

Insert Table 11.2 about here

11.2.3 The “neutral tone”

The lexical tone system of Mandarin differs from Cantonese in that some morphemes, such as the agreement-soliciting particle *-ba*, the pragmatic particles *-ma* and *-a*, the verbal suffix *-le*, and the nominal suffix *-zi*, are inherently unspecified for tone. These morphemes are said to be in the “neutral tone,” as shown in Table 11.1. (Sometimes the lack of specification is identified even more explicitly as a tone type, by calling it “tone 5”.) Containing a neutral tone syllable is one of the clearest indicators that a recurring sequence of syllables is a fully lexicalised polysyllabic word rather than a more decomposable compound or even a phrase, as in the monomorphemic *dōngxi* ‘thing’ versus the obviously compound *dōngxī* ‘east-west’. However, Mandarin varieties differ in how reliable an indicator of lexical status this is – i.e., in the proportion of polysyllabic words that do contain neutral tone. In this respect, the old Beijing dialect that Chao was describing was almost a “word accent” language when compared to modern Putonghua (Hu 1987). And Putonghua, in turn, makes greater use of neutral tone than does Guoyu (Tai 1978).

The phonetic value of the neutral tone is traditionally characterised as being “parasitic” on (or predictable from) that of the preceding full-tone syllable. However, the exact nature of this “parasitic” relationship seems to differ across varieties (and, possibly, across different phrasal contexts within some varieties). The difference between varieties is illustrated in Figures 11.1a and 11.1b, which show fundamental frequency (F0) traces for a Guoyu utterance and a Rugaohua utterance of a sentence containing the word *háizimen*. In the Rugaohua utterance, the high F0 value that is reached at the end of the Tone 2 of the initial syllable is maintained over the following two neutral-tone syllables. In the Guoyu utterance, by contrast, the F0 reaches a peak at the end of the *háí* and then falls gradually to a mid level at the end of that phrase.

Insert Figure 11.1 about here.

In Putonghua as well as Guoyu, such a fall in pitch to a mid level is typical of citation form utterances of words ending in a neutral tone after Tone 1 or Tone 2. This is illustrated in Figures 11.2a and 11.2b. When the neutral tone occurs after Tone 3, on the other hand, the fall-rise of the lexical tone is spread out over both syllables, as shown in Figure 11.2c. And similarly after Tone 4, as shown in Figure 11.2d, the fall of the full lexical tone is spread out over both syllables, so that F0 reaches a very low value at the end of the neutral tone syllable.

Insert Figure 11.2 about here

From the above, the Rugaohua pattern might be described in terms of classical “tone spreading” whereby the last tone target is copied onto syllables that surface with no tone

specification of their own (Huang 1999). However, the Guoyu and Putonghua case clearly is more complicated. The pattern in sequences with Tone 3 or Tone 4 might be captured with the kind of “spreading” that has been described for Shanghai by Jin (1985) and others. However, this will not account for the pattern in sequences with Tone 1 or Tone 2. One possible analysis is that these citation form cases show something like the “final lowering” posited for English by Liberman & Pierrehumbert (1984), among others. Perhaps there is even a L% boundary tone here (see section 11.2.6). If this analysis is borne out, then we can posit a phrase edge with a L% boundary tone after the *háizimen* in Figure 11.1a. Controlled observation of Guoyu and Putonghua neutral tone syllables in other phrasal contexts after each of the four lexical tones is needed.

In addition to having no independent tone specification, syllables with neutral tone are also characterised by segmental lenition. For example, they have shorter durations than syllables with a full lexical tone specification, and the inherently short high vowels can be devoiced after voiceless fricatives in these syllables when they follow a Tone 4 syllable, as in *dòufu* ‘beancurd, tofu’ and *yìsi* ‘meaning’ (Chao 1968:37, 141). In neutral tone syllables in Guoyu and Putonghua, unaspirated stops are typically voiced, as if foot-internal, as in *zhīdao* ‘to know’ where the /t/ in the syllable, *dao*, is [d] phonetically. One also finds contractions, such as *tām* ‘they’ alternating with the full, disyllabic form, *tāmen*, consisting of the third person singular form, *tā*, plus the *-men* plural suffix (Chao 1968:141). This connection between segmental lenition and tone status is reminiscent of the accentual system of Swedish, where “word accent” defines a categorical level of stress.

11.2.4 Stress

Thus, by contrast to Cantonese, Mandarin has stress. Moreover, it has it both inherent in the lexical entry for some morphemes (the neutral tone discussed above) and at the phrasal level. For

example, the perfective aspect marker *-le* in *xiě-le* ‘written’ and the second syllable of *kuàizi* ‘chopsticks’, unlike the initial syllables, are inherently short and reduced and not specified for tone. The word *bù* ‘not’, on the other hand, is lexically a high-falling tone, but is produced as a weak neutral-toned syllable in the A-not-A interrogative construction (e.g., *liàn bu liàn* literally ‘practise-not-practise’ = ‘Will (you) practise?’), unless the pragmatic context puts metalinguistic contrast on it.

Additionally, there is *local manipulation of pitch range* that might be called “stress” at a higher level. This is illustrated in Figures 11.3 and 11.4. The sentence in Figure 11.3a was produced with narrow focus on the subject, *Wèi Lì*, in the utterance, *Wèi Lì mài làròu* ‘Wei Li sells (Chinese) bacon,’ where the pitch range of the two focused syllables, *Wèi Lì*, is expanded, in contrast with the rest of the sentence. Such pitch range expansion does not occur in the corresponding sentence in Figure 11.3b with broad focus.

Jin’s (1996) and Xu’s (1999) data (for Putonghua speakers) suggest that these pitch range effects are not co-extensive with any phrasal unit. That is, the pitch range expansion is most obvious on the focused word, whereas the compression afterward extends over the entirety of the remainder of the phrase. Also, there can be “pauses” in the middle, as in Figure 11.4c (for a Guoyu speaker).

Insert Figure 11.3 about here.

Insert Figure 11.4 about here

Because the utterances in Figure 11.4 consist entirely of Tone 1 syllables, they also show clearly another aspect of these pitch range manipulations. When narrow focus is put on the first word in Figure 11.4b, the proper noun *Ōuyīng*, the second syllable is affected more than the first,

resulting in a rise over the word. Xu's (1999) data show a similar trend for the disyllabic noun *māomī* 'kitty' in initial position. It is tempting to relate this targeting of the second syllable to Chao's (1968) observation that sequences of full-toned syllables are not all equally prominent. That is, he describes a contrast between trochaic stress on forms such as *kuàizi*, but iambic stress on forms such *māomī*, where there is stronger stress on the second syllable.

Chao describes this stronger stress on the last full-toned syllable of a word as consistent and predictable. It occurs for both phrases and compound words (Chao 1968:360-361), for example. Thus, the stronger versus weaker stress in different positions are "all allophones of one phonemic stress" (1968:35), and the only phonemic contrast is between iambs (with two full-toned syllables) and trochees (with neutral tone on the second syllable). By contrast, Yin (1982) claims a three-way contrast among disyllabic forms with two full-toned syllables. In addition to Chao's "predictable" iambic pattern of "medium" stress followed by "heavy" stress, he cites trochaic forms with "heavy" stress on the first and "medium" stress on the second syllable, and trochaic forms with a "light" second syllable. Yin's own examples include minimal pairs for both of these types of trochaic forms. For the heavy-medium versus heavy-light cases, he gives *tèwù* 'special tasks/duties' versus *tèwu* 'secret agent, spy,' where the tone on the second syllable is frequently "neutralised" in Putonghua. And for the medium-heavy (iambic) versus heavy-medium (trochaic) cases, Yin gives *sànbù* 'to take a walk' versus *sànbù* 'to scatter.' Our attempts to elicit native speaker judgments do not support either Chao's claim or Yin's. That is, most native speakers of standard Mandarin deny any differences in stress level beyond the palpable contrast between (all) full-toned syllables and the neutral-tone syllables. However, stress is notoriously difficult to access with introspective judgments, and we need to look for evidence in patterns of production, such as potential contrasts

in the domain of the pitch expansion under narrow focus, as in the particular targeting of the second syllable that is depicted in Figures 11.3 and 11.4.

Another place to look for evidence of a contrast between “strong” (or “primary”) stress and “medium” (or “secondary”) stress on full-toned syllables is in the possibility (and likelihood) of tone “neutralisation” in running speech, as in the *bu* of *liàn bu liàn* cited above. Just as varieties differ in the distribution of neutral tones in the lexicon, they also differ in regards to where neutral tone can occur in sentences. The A-not-A construction is a good illustration both of the effect and of the differences between Putonghua and Guoyu. The *liàn bu liàn* example cited above is Putonghua, and the second repetition of the verb is optionally also in neutral tone (Chao 1968:270), making a ternary foot. The full form of the verb in a short declarative (as in the answer ‘I’ll practise.’) would be *liànxí*, or in connected speech *liànxi*, a more normal binary foot. A (rather pedantic) alternative to *liàn bu liàn* would be to repeat the whole verb: *liànxi bu liànxi*. In this case, the second *liàn* could not be in neutral tone. In Guoyu, the rhythmic constraints differ, and so does the syntax. The verb *liànxí* does not alternate readily between full and neutral tone on the second syllable, but is almost always produced with an identifiably rising Tone 2 on the second syllable (albeit somewhat reduced relative to the pitch fall on the first syllable). Also, the A-not-A construction incorporates the partial reduplication of the homologous Taiwanese construction so that the normal way to ask ‘Are you going to practise?’ is *liàn bu liànxí*, with neutral tone only on *bu*. That is, the second *liàn* is foot-initial and never reduced in Taiwan Mandarin.

As we have already suggested, the trochaic words with neutral tone can be described in terms of the stress foot. Neutral toned syllables cannot occur in isolation, and there are no content words that have only neutral-tone syllables. In many descriptions in the framework of Autosegmental Phonology and Metrical Phonology, this is accounted for by saying that a full-toned

syllable can be a foot on its own, whereas a neutral-tone syllable is necessarily footed together with the preceding full-toned syllable, making the prosodic word in Mandarin minimally a bimoraic foot (see Yip 1980, Wright 1983, Duanmu 1990, inter alia). Shih (1986, 1997) proposes a somewhat different definition of the foot, in which even a full-toned monosyllabic form does not form a foot except under exceptional circumstances. (Beattie's, 1985, definition of the "prosodic word" is similar, as is Chen's, 2000, "minimal rhythmic unit.") Shih's foot thus emphasises the predominately bisyllabic rhythm of the language, a metrical organisation that is reflected the lexical statistics (i.e. the much higher prevalence of disyllabic words in Mandarin than in Cantonese) as well as in some accounts of third tone sandhi, covered in the next section.

11.2.5 Tone sandhi and the "superfoot"

Some of the lexically specified tones are subject to tone sandhi conditioned by tonal context and some other prosodic factors. The facts are clearest for Rugaohua. In this variety, the tone sandhi processes are similar to those found in the neighboring Wu dialects such as Shanghainese (see Jin 1985, Selkirk & Shen 1990, Duanmu 1997, Chen 2000) in that only the first syllable in a sandhi domain keeps its original tonal specification while a rightward spreading forces the other syllables in the word or phrase to lose their underlying tones. Just as in Shanghainese, the sandhi domain in Rugaohua is constrained by such factors as syntactic structure, focus, speech rate, and so on. Some examples are shown in (1) below:

- (1) a. /da⁴¹ + ɕiaŋ³⁵ + sã⁴¹/ → [da⁴¹ ɕia(ŋ)¹¹ sã⁴¹] 'college student'
 b. /ɕio³²³ + ɕiaŋ³⁵ + sã⁴¹/ → [ɕio³² ɕia(ŋ)³³ sã⁴¹] 'elementary school pupil'

Each of the above three-syllable utterances has two sandhi groups, so that only the first and third syllables keep their underlying lexical tones. This corresponds to the left-branching syntactic structure shown in (2):

(2) [[X Y] Z]

The same three-morpheme sequences could be uttered with the first and second syllables bearing their lexical tones, and the third syllable grouped together with the second, as in (3), but then the syntax is right-branching, and the interpretation different.

(3) a. /da⁴¹ + ɕiaŋ³⁵ + sǝ̃⁴¹/ → [da⁴¹ ɕia(ŋ)³⁵ sǝ̃⁵⁵] ‘big student’

b. /ɕio³²³ + ɕiaŋ³⁵ + sǝ̃⁴¹/ → [ɕio³²³ ɕia(ŋ)³⁵ sǝ̃⁵⁵] ‘little student’

Stress patterns change accordingly. In a phrase with the sandhi grouping pattern of (1), syllables X and Z are stressed while Y is relatively unstressed, whereas in the pattern of (3), syllable Y is stressed and Z is relatively unstressed. Whether this is secondary stress or a total reduction to the level of a neutral tone syllable remains a question for further investigation.

The standard varieties of Mandarin also have tone sandhi. Most notably, Tone 3 changes to Tone 2 when followed by another Tone 3, and the occurrence of this “third tone sandhi” seems to bear some relationship to stress, or at least to the bisyllabic rhythm mentioned above. However, there is much that is still not understood about the phonetics of third tone sandhi, despite much discussion on the topic in the literature (e.g., Chao 1968, Cheng 1973, Shih 1986, Kratochvil 1987,

Zhang 1988, Hung 1989, Shen 1990, Hsiao 1991, Chen 2000). Thus, Cheng (1973) emphasises the blocking of third tone sandhi at certain syntactic (or prosodic) boundaries (reminiscent of the facts just described for Rugaohua), whereas Shih (1986, 1997), Hsiao (1991), and Chen (2000) point to rhythmic constraints, and posit a “superfoot”—i.e. a grouping of a stranded monosyllable together with the adjacent bisyllabic unit. (Shih (1997) reviews other prosodic factors as well, such as the “refooting” of syllables that might otherwise be foot-final under conditions of prosodic prominence for narrow focus. And she points to “morphological” factors such as sequence frequency.) Even the categorical nature of the sandhi change is disputed. For example, Wang & Li (1967) and Peng (1996) show that the sandhi tone is indistinguishable from underlying Tone 2 in identification tests, although Zee (1980) and Peng (1996) both show “incomplete neutralisation” in controlled instrumental studies. It is also possible that the facts differ across the major standard varieties.

Chao (1968) also describes various “phonetic” tone sandhi changes. For example, Tone 2 in a trisyllabic sequence before a full-toned syllable changes into Tone 1 when the first syllable is Tone 1 or Tone 2. Underlying *cōngyóubǐng* ‘onion oil cake,’ for example, is pronounced *cōngyōubǐng* in styles of Putonghua that are close to Beijinghua. There is also “morphophonemic” tone sandhi for the word *bù* ‘not’ and the Tone 1 numerals *yī* ‘one’, *qī* ‘seven’, and *bā* ‘eight’. For example, *bù* becomes Tone 2 before a Tone 4 morpheme—e.g., *bùgǒu* ‘not lax’ versus *búgòu* ‘not enough’. These other sandhi phenomena and their relationship to stress and/or grouping are even less well understood than is third tone sandhi. Thus, much work is needed before we can say definitively whether or not tone sandhi phenomena define a level of grouping comparable to the clear sense of a “phonological phrase” that the Rugaohua tone sandhi imparts.

11.2.6 Evidence for higher-level prosodic grouping

In addition to lexically specified tones, sometimes there are also pragmatic tones at the ends of sentences. Chao (1968:812) posits a Rising Ending and a Falling Ending for Beijinghua, and analyses them as “particles” on phrases and sentences. That is, he describes them as localised to the last syllable (specifically, to the voiced portion of the last syllable) of an utterance, revising an earlier analysis (Chao 1933) in which they are treated as part of a more global backdrop intonation contour. If Chao’s observation is correct, his “Ending” patterns would be tagged as boundary tones in a ToBI framework system.

Chao’s observations were of Beijinghua from before the emergence of a separate standard in Taiwan. In our observations to date, we have identified at least two boundary tones also for Guoyu and Putonghua, which we tentatively identify with Chao’s Rising Ending and Falling Ending (although unlike Chao, we have not observed them yet on syllables other than sentence-final particles, which are lexically unspecified for tone). One has a higher tonal target than the other. In Figure 11.5a, *tāmen bú mài yúsǎn ma?* ‘Don’t they sell umbrellas?’, the sentence-final particle *ma* is produced with a high boundary tone. The sentence in Figure 11.5b, on the other hand, ends with a low boundary tone. In the first case, the speaker is asking a yes-no question, but the boundary tone suggests a presupposition that the store should sell umbrellas. Thus, this can convey surprise, if the addressee is someone who was sent to buy an umbrella and came back empty-handed. In the second case, by contrast, the L% boundary tone effectively makes the utterance a statement. It might be produced by a speaker to soften an explanation of why he came back empty-handed. The English equivalent might be something like, “Well, but they don’t sell umbrellas.”

Insert Figure 11.5 about here

In addition to these boundary tones, there are also global pitch range effects that can signal contrasting pragmatic meanings. For example, Figure 11.4a is a statement, *Ōuyīng mō māomī*, meaning ‘Ouying strokes kitty.’ When the utterance is produced as an echo question, the overall pitch range is raised, as shown in Figure 11.4d. In older impressionistic descriptions such as Chao (1933), there was no differentiation between these global “intonations” and the more localised boundary tone effects described above. As already noted, Chao (1968) implicitly recognises the difference, but does so by describing only local effects, and Kratochvil’s (1998) account is similar. However, some recent instrumental research, such as Shen (1990), has ignored the local effects to focus on the interaction of sentential pitch range effects with the lexical tone specification in short simple sentences such as the utterances in Figures 11.4a and 11.4d. Longer utterances in more varied pragmatic contexts than those provided by Shen seem to show both global and local effects, interacting in ways that seem quite complicated, given our current limited understanding of the levels of prosodic grouping that are available to be the domain of phrasal pitch range specification. For example, some of our observations of Putonghua speakers suggest a three-way distinction among sentences ending with *ma*. There is a “higher than neutral” ending, with a raising that does not start at the beginning of the sentence, but which also does not seem to be confined to the last syllable. And this pattern contrasts both with a “neutral” final lowering and with a “lower than neutral” ending (see Lee 2000 for a discussion of the phonetics and pragmatics of these effects). Substantial new studies are necessary to identify the domains of these effects, and to specify their relationship to the domain of the boundary tones.

These studies also should control carefully for factors such as prominence, which Jin (1996), Xu (1999), and others have also shown to affect pitch range (see section 11.2.4). They also

should investigate the relationship to the domain of tone sandhi in varieties such as Rugaohua (see section 11.2.5). While any definitive statement is premature, it seems safe to speculate that the boundary tones and global pitch range effects can be identified with a larger prosodic grouping that is comparable to the “intonation phrase” of Cantonese.

11.3 Mandarin ToBI

Although our current understanding is not complete enough to propose a definitive set of annotation conventions for any variety of Mandarin, developing a preliminary set of tags for the phenomena described in section 11.2 helps to clarify the outstanding questions for further research. A preliminary codification also provides a useful tool for investigating many of these phenomena using existing spoken language corpora. Since Mandarin is an economically and politically important language, it comes as no surprise to find that there are both proprietary and publicly available corpora. For example, the Linguistic Data Consortium has databases for both the PRC and Taiwan standard varieties. It is also not surprising that there have been several independent efforts to develop systems for tagging Mandarin corpora within the ToBI framework.

In this section, we will describe the union of two of these systems, one developed at Academia Sinica (AS) in Taipei and the other at Ohio State University (OSU) in Columbus. In preparing for the satellite workshop at ICPHS’99, these two sites agreed to merge their systems into a single set of conventions that eventually might be applied to utterances, both spoken and read, in any variety of Mandarin — i.e. a Pan-Mandarin M_ToBI system. The two original systems were developed for rather different purposes. The AS conventions were developed for purposes such as developing the prosody component of a Guoyu text to speech (TTS) synthesis system, and training and testing automatic speech recognition (ASR) models using read speech. (It has also been

adopted by the Institute of Linguistics at the Chinese Academy of Social Sciences in Beijing, to develop a Putonghua TTS system — see Li, Zu, & Li 1999.) The OSU conventions were developed for purposes such as exploring the relationship between prosody and pragmatics using spoken language data, and documenting prosodic variables for sociolinguistic models of linguistically heterogeneous speech communities, including but not limited to Taiwan. Thus, the merging of the two systems is in keeping with the spirit of the original ToBI framework system for English. We are trying to forge a communal standard that can be shared across a viably large and varied community of users.

At this writing, the development of the merged system is not complete. Determining the final set of M_ToBI labels for Guoyu or Putonghua alone will require several more iterations of the same process that went into the development of the original AmEng_ToBI. (That is, participating sites contribute to a pooled set of calibration utterances that everyone transcribes. We calculate inter-transcriber agreement and discuss major discrepancies. We agree on a manageable set of changes to accommodate the problems that have emerged. The next iteration then tests the modified system against a new set of calibration utterances.) And codifying the full Pan-Mandarin system will be an even longer-term process, since our understanding of intonational phenomena in even the major regional varieties is very sketchy by comparison to research on either of these two standard varieties.

Even in its current unfinished state, however, the merging of the systems has already served to highlight several of the most salient prosodic phenomena of the language — particularly, phenomena that were being treated similarly in the two original systems. Most noteworthy among these is the set of phenomena described above in section 11.2.4. Both sites had independently chosen to tag syllable prominence using an explicit set of hierarchical labels (for “emphasis” or

“stress” levels) that is separate from the tags for words, tones, and break indices common to all ToBI framework systems. A third system, developed at Lucent Bell Laboratories, is also for a Guoyu database, and it also tags stress levels. One of the first things that we hope to do in developing our Pan-Mandarin ToBI further is to conduct a three-site calibration experiment, in which each of us transcribes representative utterances from the other two sites, to develop mappings between these tags across all three systems.

With the caveat, then, that many details almost certainly will have changed in the final standard, we describe in this section the symbolic tiers of the current state of the merged M_ToBI system. We will relate these tiers to the phenomena described above in section 11.2. We also will identify other ways in which the two systems either were in accord or complemented each other enough to make it possible to conflate the two sets of tagging conventions. Eight tiers are proposed at this stage in the system. These tiers are listed in Table 11.3, and discussed in turn in the remainder of this section: the words and romanisation tiers in section 11.3.1, the syllables, stress, and sandhi tiers in section 11.3.2, the tones tier in section 11.3.3, the break indices tier in section 11.3.4, and the code tier in section 11.3.5.

Insert Table 11.3 about here

11.3.1 Transcribing the “words”

As in other ToBI framework systems, the symbolic tags in an M_ToBI transcription are intended to be anchored in time to an audio recording of the utterance. In ToBI framework systems for most other languages, the initial set of time stamps is via an obligatory **words** tier. The words tier provides a label for each word in the utterance in the native orthography (or in a romanised transliteration, in the case of Gr_ToBI — see Arvaniti & Balthazani, this volume). It thus orients

the transcriber to the signal and accompanying F0 contour in terms of a familiar representation that is accessible to anyone who knows the language. It also promotes the development of tools for automating aspects of the transcription using resources such as online pronunciation dictionaries.

In keeping with these functions, the words tier for Mandarin is defined as a syllable-by-syllable transcription in Chinese characters. Although Mandarin has far fewer monosyllabic words than Cantonese, having a separate tag for each syllable accords with native speaker intuitions about what “words” are in Chinese. Moreover, the use of Chinese characters provides a common set of easily accessible tags that “normalise” over the substantial segmental differences among regional varieties. The OSU group has implemented M_ToBI in Emu (Cassidy & Harrington 2001), a database labelling/querying language that allows Chinese character input on a PC/Windows platform using standard encodings such as Unicode and Big5. However, the labelling platform that is common to both original sites (the UNIX/Linux based *xwaves/xlabel* program) does not allow Chinese character input. Therefore, M_ToBI currently does not specify the words tier as obligatory. That is, we allow the orthographic transcription to be stored in a separate text file, just so long as it can be linked indirectly to the audio file via some other tier that also provides a syllable-by-syllable transcription of the utterance.

In the current M_ToBI system, this other tier is by default the **romanisation** tier (abbreviated as *romazi* in the label file extension). The romanisation tier tags each orthographic “syllable” using an “ASCIIified” version of the Pinyin system. This modified Pinyin romanisation is strictly ASCII in that lexical tones are marked using Chao’s tone numbers rather than with the original Pinyin tone diacritics. For example, Figures 11.5a and 11.5b show the lexical tone on all syllables except the plural marker, *men*, and the sentence-final particle, *ma*, both of which are in the neutral tone. A further modification from Pinyin is that each syllable is labelled with the

“dictionary form” for the particular variety in which the utterance is produced. That is, we have decided not to “normalise” to a Putonghua reading for the Chinese character that would be used to transcribe the syllable in the associated text file (or the currently optional words tier). Some of the ramifications of this decision are illustrated in Figure 11.6.

Insert Figure 11.6 about here

The Pinyin romanisation system was originally developed for Putonghua. We know that it is adequate for transcribing the other two national standards, Guoyu and Huayu, because they are segmentally very close to Putonghua. Pinyin also provides some “extra” characters to transcribe sounds found only in the prestigious Beijing regional variety. (Standard broadcast Putonghua is sometimes equated with Beijing Mandarin, but the two are distinct — see, e.g., Hu 1987, Chan & Tai 1989.) Other regional varieties require further additions, such as the ‘K’ that we have added to transcribe the final glottal stop in the Rugaohua word, *qi@K⁵* (/tɕiəŋ⁵/) ‘to eat’ in Figure 11.6.

Before we can codify the romanisation tier in the final M_ToBI system, we need to test how well Pinyin does with other regional varieties. If too many added characters are needed to cover the full range of segmental variation that is observed, it may be more tractable to replace Pinyin with something closer to a phonetic transcription. In that case, we will adopt instead the SAMPA-like transcription that the Academia Sinica group developed for segmental transcription.

While it is not part of the original ToBI framework system for prosodic transcription, the Academia Sinica (AS) Guoyu databases are all labelled with this phonemic alphabetic transcription of “initial” (onset) and “final” (rhyme) segments (see Tseng & Chou 1999). The transcription standard was designed to be compatible also with segmental transcriptions for Taiwanese and

Hakka (Kejia), and it should extend easily to regional varieties of Mandarin, too. For standard Mandarin utterances that are transcribed at Academia Sinica, these segmental tags can be aligned automatically with the signal from the orthographic transcription file using the text preprocessing component of the site's TTS system in combination with an HMM recogniser (see Chou, Tseng & Lee 1998).

11.3.2 Transcribing syllable prominence and tone sandhi

As noted above in section 11.2.3, orthographic syllables do not always correspond to phonological syllables, because of contractions as such as *tām* for *tāmen* 'they'. The syllables tier allows us to capture such discrepancies, by labelling only the phonological syllables. In its current form, these syllables are transcribed in a broad phonetic transcription using the same Pinyin-based symbols as in the romanisation tier. As more spontaneous speech is recorded and transcribed in the M_ToBI system, we may find it useful to differentiate this tier even more from the orthographic tier, by recording major allophones.

In the Emu implementation of the M_ToBI standard, the syllables tier (shortened to 'Syllable' in the figures here) has two other labelling fields that mark the relative degree of stress of each syllable in the utterance (stress tier) and any tone sandhi (sandhi tier). In the xwaves/xlabel implementation, the relationship among these three tiers must be insured by an independent grammar checker. We describe the stress and sandhi tiers in turn below.

The general idea of transcribing stress levels was common to both original systems. The two systems agreed even on the number of distinct levels to transcribe. In the AS system, the four labels E0 through E3 were defined in terms of "reduced" versus "normal" versus "moderate" versus "strong" levels of "emphasis" for each syllable. However, by comparison to the development of

the break index definitions (see section 11.3.4), little attention had been given to examining how these definitions were understood by different transcribers. Thus, we need a thorough inter-site calibration experiment before we can be completely sure how the AS E0 through E3 map onto S0 through S3 at the other site. In the interim, the preliminary M_ToBI standard has adopted the more specific definitions of the four “stress” levels in the OSU system. These levels are summarised in Table 11.4.

Insert Table 11.4 about here

Stressed syllables with fully realised tone are labelled with **S3**. This level is illustrated in each of the utterances in Figure 11.5 by the first syllable of the subject *nǐmen* ‘you (plural)’ and all three syllables of the verb phrase *mài yúsǎn* ‘sell umbrellas’. Syllables with lexical neutral tone are inherently unstressed and, therefore, are labelled with **S0**. This level is illustrated twice in each of the utterances in Figure 11.5, where both the plural marker *-men* of *nǐmen* ‘you (plural)’ and the sentence-final particle *ma* are labelled with **S0**. In running speech, some syllables with lexical full tones are often “neutralised”. That is, they are produced with the tonal and temporal characteristics of lexical neutral tones. In Figure 11.7, in the A-not-A construction *liàn bú liàn*, the last two syllables, *bú liàn*, were neutralised. Both syllables have “lost” their tonal specification. That is, there are no traces of a high-rising tone on *bú* or of a high-falling tone on *liàn*. Lacking definitive evidence that this tone “loss” makes such syllables “feel” as reduced as an “inherently” neutral tone S0 syllable, these syllables are labelled with **S1**. **S2** is then used to label syllables with substantial tonal reduction that stops short of this complete “neutralisation” of tonal specification. That is, S2 labels a substantial undershoot of the tonal target, which is usually accompanied by at least some

shortening of duration. For instance, in Figure 11.8 the high-falling tone *kuài* does not reach the very low tonal target suggested by the ‘1’ in the tone transcription.

Insert Figure 11.7 about here.

Insert Figure 11.8 about here.

Note that we have characterised this fall to mid level in terms of a phonetically gradient undershoot of the tone target that is tagged only implicitly by the S2 label. By contrast, Chao (1948:26) described it in terms of a tone sandhi alternation between a “full” falling tone /51/ and a “half” falling tone /53/. Later, Chao (1968:28-29) accounted for the “full” versus “half” falling tone as the effect of stress, with greater stress enlarging both the range and the length of the tone. And since he analyses disyllabic words with full-tone on both syllables as having iambic stress, it is the first syllable that is realised with the partial falling tone. In this later formulation, Chao considered such stress-related phonetic differences to hold true for any combination of tones, and did not limit it to Tone 4. Thus, in both formulations (albeit more clearly in the later one), Chao distinguished this kind of “phonetic” sandhi from the clearly categorical alternation involved in the “third tone sandhi” that Taiwanese school children are taught in elementary school. He described the undershoot in terms of the alphabetic alternation only because he did not have access to a more gradient phonetic representation such as the F0 contour that is a necessary part of any ToBI framework transcription.

In the current M_ToBI system, then, we adopt Chao’s distinction between “phonetic” and “phonological” sandhi. We use S2 alone to tag the variation in tone shape that arises from undershoot and coarticulation, but provide an explicit alphabetic transcription for variation such as

third tone sandhi where there is clearer evidence that a categorical alternation is involved. (This follows the practice of the OSU system. The AS system circumvented the issue by not transcribing tone at all. Instead, the TTS prosody component was trained on the AS read-speech database to extract a “tone pattern” that could be generated for each different sequence of tones that can occur within a minor prosodic phrase — see section 11.3.4.) For categorical alternations, then, the sandhi form of the affected tone is marked on the sandhi tier with transcription in Chao’s tone numbers. Thus, **35** represents the third tone sandhi of the standard varieties, where Tone 3 is phonetically realised as a rising tone. This is illustrated in Figure 11.9, where the first, the third, and the fourth syllables of the sentence were produced with the **35** sandhi tone. (Figure 11.6 shows analogous examples for sandhi forms in Rugaohua.)

Insert Figure 11.9 here.

The tone sandhi tier is also used to label two other types of alternation that are not always termed “sandhi” in the literature. For example, **214** is used to label Tone 3 produced as a low dipping tone in any of the standard varieties. As noted above in section 11.2.2, this is an “allotone” of Tone 3 which may occur in sentence- or phrase-final position, or when the syllable is produced with emphatic prominence. It is illustrated by the last syllable in Figure 11.9. We can think of this alternation as position-specific tone sandhi. The other type of alternation is morpheme-specific tone sandhi. For example, the negative morpheme *bù* in the standard varieties has an alternate rising form when it is followed by another high-falling tone, as in *bú yào* ‘don’t want’. Such morpheme-specific variation is also transcribed on the sandhi tier, as shown by the rising alternate of *bù*, marked with **35** in Figure 11.7.

Note, however, there are alternative analyses for both cases. For example, one traditional description of the /21/ alternate of Tone 3 is that it is a phonetic “truncation” of a more canonical /214/ in prosodic positions where syllable duration is short for other reasons, such as not being phrase-final (although see section 11.2.2 for difficulties with this analysis for Guoyu). In a similar way, since the negative morpheme *bù* ‘not’ is typically subject to “neutralisation” in running speech, it is possible to think of the *bú* alternate as a lexicalisation of the “neutralised” variant in an unusual “upbeat” position, where the tone shape cannot be “parasitic” to the preceding tone. That is, the rise might be described as a “phonetic interpolation” to the high target at the beginning of the *yào*. If the “upbeat” analysis is correct, then the 35 tagging on the sandhi tier for the rising alternate of *bù* should be replaced by an S1 specification on the stress tier. If the categorical sandhi analysis is correct, on the other hand, then the 35 tagging of *bù* on the sandhi tier should always be accompanied by stress level S2 or higher on the stress tier. In the current M_ToBI system, we do not enforce the sandhi analysis, but instead allow transcribers to combine **S1** with the 35 tag for *bù*.

As mentioned above in sections 11.2.4-11.2.5, the relationship between stress and tone sandhi is a very under-researched area. It is also likely that the relationship is different for different varieties, and the conflicting analyses of the rising variant here may reflect real differences between the native varieties of the linguists proposing them. Explicitly tagging the alternation on an independent sandhi tier without requiring the stress level to be higher than S1 prompts us to explore these possibilities in a way that we could not do if we codified the system prematurely. Moreover, the tone sandhi patterns for different regional varieties are as varied as the segments and lexical tones. Having a sandhi tier independent of the more canonical tones tier (described in the next section) encourages researchers to develop labels that are specific to each new variety that is

transcribed, without making a premature commitment to the relationship between any sandhi pattern and the intonational phrasing.

11.3.3 Transcribing the “intonation”

Because the distinction between tonal reduction and tone sandhi is still an open issue, and because the evidence for boundary tones is less clear than in Cantonese, the current version of M_ToBI does not follow the C_ToBI convention of tagging lexical tones and boundary tones together on the same tier. Instead, we transcribe the “underlying” lexical tone on the romanisation tier, the potentially more rhythm related effects of tone sandhi on the sandhi tier, and make a third tier for the boundary tones and pitch range effects discussed in section 11.2.6. That is, the M_ToBI version of the usual ToBI **tones** tier includes only these “true” intonational effects.

The tones tier marks global or local characteristics of the backdrop pitch range, such as a general downtrend versus raised pitch for a sentence, as well as the phenomena that were described in terms of boundary tones in section 11.2.6. (The AS system circumvented the issue of whether these “true edge events” can be distinguished from more global trends by training the TTS system to generate a stylised “intonation contour” for each breath group, based on the punctuation of the read text — see Chou, Tseng & Lee 1996. In order to generalise to spontaneous speech and to more different read-speech styles, we included the “intonation” tags of the OSU system in the merged M_ToBI standard. This is an example of a way in which the two systems complemented each other.)

Insert Table 11.5 about here

Tags for boundary tones are like those in other ToBI framework systems. Thus, as shown in the set of tags for the tones tier in Table 11.5, **H%** and **L%** mark high boundary tone and low boundary tone, respectively. (See Figures 11.5a & 11.5b (where the tones tier is given as ‘Tone’), and the discussion above in section 11.2.6.)

Tags for backdrop pitch range effects are descriptive terms set off with the same “%” that identifies the boundary tones. For Putonghua and Guoyu, there are two types of pitch range effects that are tagged in this way: ones that describe a backdrop “contour” for an entire sentence or phrase (see, e.g., Shen 1989, for Putonghua), and ones that are associated with more localised effects of focal prominence (see, e.g., Jin 1996, Xu 1999). However, the syntax is the same. The tag is placed at the beginning of the effect, which is also the beginning of the relevant phrasal unit in the first type of tag. Thus, **%reset** marks the beginning of a new pitch downtrend over a “normal” declarative phrase (i.e., a pitch reset), whereas **%q-raise** marks the flat raised pitch range regularly seen in echo questions. As shown in Figure 11.3b, the %reset label is placed at the beginning of the phrase, to signify the beginning of the pitch downtrend. Figure 11.3c shows the corresponding syntactically unmarked echo question, and the sentence-initial %q-raise tag. (Another pair of examples is the statement in Figure 11.4a and the corresponding echo question in Figure 11.4d.)

The placement of the second type of pitch-range tag is illustrated in Figure 11.3a. The beginning of a local expansion of pitch range due to emphatic prominence is marked with the tag **%e-prom**. The expansion is often accompanied by reduction of the pitch range of the following syllables, which is signaled by **%compressed** at the onset of pitch range reduction. The sentence in Figure 11.3a was produced with narrow focus on the subject noun, *Wèi Lì*. The beginning of the expanded pitch range over these two emphasised syllables is marked with %e-prom, and the

beginning of the compression of pitch range on the succeeding syllables is marked with %compressed.

The tags described here reflect our current understanding of the standard varieties of Mandarin. The tones tier conventions for Rugaohua probably will have to be different. For example, our casual observation to date suggests that tags such as %e-prom can be linked to break indices for phrasal units at the level of the tone sandhi group, as in the neighboring Wu varieties (see, e.g., Selkirk & Shen 1990, and the discussion in section 11.2.5). However, there is as yet no extended study to confirm this. Moreover, our knowledge of such intonational phenomena in other regional varieties is even sketchier. We should not assume that tones tier labels that are appropriate for Putonghua can be applied to any regional variety except as an initial working hypothesis in the development of variety-specific conventions.

11.3.4 The break indices

As in other ToBI framework systems, prosodic phrasing is represented with a hierarchy of break indices, with a number (a particular level of disjuncture) tagged at the end of every segment on the words tier and/or romanisation tier. The merged M_ToBI system adopts the break-index conventions of the AS system which have been tested against the entire Guoyu TTS database, in an inter-transcriber calibration experiment involving the two primary transcribers (see Tseng & Chou 1999). The conventions are summarised in Table 11.6.

Insert Table 11.6 about here

Six levels of prosodic juncture are distinguished. Level 4 marks boundaries which correspond to a reset of pitch between sentences or phrases. If a breath group boundary is accompanied by a prolonged pause, it is labelled with break index 5. The first of these two labels (breath group with only a small pause) is illustrated by the utterance fragment in Figure 11.10, which is part of a long sentence produced as two breath groups. There is a large reset accompanied only by a short pause after the verb, *shuō* ‘to say’, which sets off the quoted material from the matrix clause. In the AS calibration experiment, a large percentage of inter-transcriber disagreements involved confusion between B4 and B5. Even after exchanging notes and retranscribing, 21% of the breaks that were labelled with B4 or B5 by either transcriber were labelled with the other of these two tags by the other transcriber. This is unsurprising for a distinction which is not a categorical one, but a matter of gradient differences in the degree of final lowering before the boundary, the amount of pitch range reset after the boundary, and so on. In the final M_ToBI system, these two labels may be collapsed and the distinctions captured instead in a direct metric of pitch range that can be correlated with a recursive model of discourse structure. (Cf. also the “finality tier” in the J_ToBI conventions described by Venditti, this volume.)

Insert Figure 11.10 about here.

Break indices **2** and **3** are used for phrase boundaries within breath groups. Here, the M_ToBI conventions distinguish between major phrases and minor phrases. The two transcribers in the AS calibration experiment were instructed to transcribe B3 when they perceived a pause, and B2 when no pause was perceived. These two labels were next most confusable after B4 versus B5. Even after comparing notes and retranscribing, there was 9% confusion between these two tags.

Figure 11.10 shows two examples of break index 2 and one example of break index 3, as identified by transcribers at the OSU site who were given the same instructions as the two transcribers in the AS calibration experiment. It is possible that further inter-site calibration tests will uncover better criteria for defining major versus minor phrases. Even if no categorical markers are identified for the standard varieties, however, these two levels of grouping probably should not be collapsed, since B2 probably will be needed for the tone sandhi group in Rugaohua.

Break index **1** is the “default” boundary between syllables, and is left unmarked in the figures, to avoid clutter. (In transcribing with the xwaves/xlabel platform, these can be inserted automatically and then changed for other break index levels.)

Break index **0** tags boundaries that have been “deleted” when a syllable is extremely reduced (S0 or S1) and its vestige segments re-syllabified with the previous syllable or the following syllable. For example, in Figure 11.8, the second syllable of *nǐmen* ‘you (plural)’ is reduced to just the initial bilabial nasal and re-syllabified with *nǐ*, so the boundary between *nǐ* and the plural suffix, *-men*, is labelled with B0.

11.3.5 The code tier

As noted above, the M_ToBI conventions are a “Pan-Mandarin” system; they are intended to be applicable to utterances in any variety of Mandarin. The conventions also are intended to serve as a tool for exploring prosodic variables in complex speech communities where speakers typically are fluent in several varieties of Mandarin (or in Mandarin and a non-Mandarin variety of Chinese), and can switch easily between codes for sociolinguistic or stylistic effect. We have noted at several places in the discussion above that the current inventory of labels may need to be expanded or modified to cover other varieties besides Guoyu and Putonghua. Rather than “recompile” the

inventory of labels each time M_ToBI is extended to a new regional variety, it would be convenient to add labels or provide variety-specific interpretations of old labels, as appropriate for the system. This more modular approach, however, requires that we identify the variety being transcribed so that labels that are similar across varieties can be interpreted accordingly. (See Bruce, this volume, for a discussion of regional variation in the phonetics of the accent 1 versus accent 2 contrast in Swedish. See also Grice et al., this volume, for the more extensive variation that needs to be covered in developing a Pan Italian ToBI.) Ideally, the mechanism that identifies the variety can then also serve to identify points where the speaker incorporates features of another variety in prosodic code-switching.

The **code** tier is being developed for these two related purposes. Labels on this tier identify the variety of Mandarin used to produce the utterance and also mark points of code-switching between varieties. The syntax of these labels is similar to that of tags on the miscellaneous tier in English. Each tag consists of a start marker “<” followed by an abbreviation for the variety. Thus <GY, <PTH, and <RGH represent Guoyu, Putonghua, and Rugaohua, respectively. The label is placed at the point where the speaker starts to use the variety in question.

More complex labels can also be created to identify the particular feature(s) involved in a partial code-switch. This is illustrated in Figure 11.1b. The utterance is a reading in Rugaohua of a text originally composed in Guoyu. Most of the lexical items have everyday cognates in Rugaohua, but *háizimen* is a word “borrowed” from the standard variety of Mandarin (which for the Rugaohua speaker would be Putonghua rather than Guoyu). Since she was reading a prepared text, the speaker produced the Putonghua word *háizimen* using Rugaohua tones, rather than substituting the native Rugaohua word for ‘children’. On the code tier in the figure, the sentence-initial <RGH marks the variety of Mandarin used to produce the utterance. The <PTH-word label

then marks the beginning of the standard Mandarin lexical item, and <RGH signals the point where the speaker resumes a pure Rugaohua code.

11.4 Conclusion.

In this chapter, we have described some of the salient prosodic phenomena of Mandarin, and proposed the development of a Pan-Mandarin M_ToBI system. Much research will be required before we have a complete system that can cover all three national standards and at least a few of the major regional varieties such as those spoken in Rugao, Xi'an, Chengdu, Nanjing, and so forth. It probably will be several years before we can even say with confidence that the broad outline of the overall structure of tiers is adequate to cover all varieties. Mandarin is spoken over a vast geographical area. Different regional dialects abut other language varieties that range from Shanghai (a Wu variety of Chinese) in the east to Tibetan (a Tibeto-Burman language) and Mongolian (an Altaic language) in the extreme west and north of the Mandarin-speaking region of the PRC. The range of prosodic variability probably is much larger than that for Cantonese (see Wong et al., this volume).

However, even at this preliminary stage, it is clear that developing a Pan-Mandarin M_ToBI system gives us a manageable way to frame the questions for future research. For example, the separation of degrees of perceived juncture (on the break indices tier) from phrasal pitch range effects (on the tones tier) allows us to explore the first of these in inter-transcriber calibration tests without committing prematurely to categorical definitions of prosodic units in a strictly layered hierarchy. The separation gives us a theory-neutral way to explore the potentially quite variable relationship across different styles and different varieties between prosodic grouping or stress, on the one hand, and tonal reduction versus tone sandhi, on the other.

It is also clear that the ToBI framework will need to be extended in at least two ways to apply to Mandarin. First, unless we limit the scope of M_ToBI to Guoyu (or perhaps to just the three national standards), we will need a device such as the code tier to accommodate to prosodic differences across varieties and to describe phenomena such as tonal code switching. Second, because Mandarin is a “stress” language, and because even the standard varieties differ markedly on the distribution of unstressed syllables relative to words and phrases, stress needs to be tagged explicitly. For Mandarin then, the ToBI framework probably should become the ToBISL framework (“Tones, Break Indices, and Stress Levels”).

References

- Arvaniti, A. and Balthazani, M. (this volume), ‘**Intonational Analysis and Prosodic Annotation of Greek Spoken Corpora**’.
- Beattie, D. A. (1985), ‘Mandarin Tone Sandhi and the Interaction of Phonology and Syntax’, Unpublished M.A. thesis, University of Toronto, Toronto.
- Bruce, G. (this volume), ‘**Intonational Prominence in Varieties of Swedish Revisited**’.
- Cassidy, S. and Harrington, J. (2001), ‘Multi-level Annotation in the Emu Speech Database Management System’, *Speech Communication* 33.1-2: 61-77.
- Chan, M. K.-M. (1991), ‘Contour-tone Spreading and Tone Sandhi in Danyang Chinese’, *Phonology* 8: 237-59.
- _____ and Ren, H.-M. (1989), ‘Wuxi Tone Sandhi: from Last to First Syllable Hominance’, *Acta Linguistica Hafniensia* 21.2: 35-64.
- _____ and Tai, J. H.-Y. (1989), ‘A Critical Review of Norman’s *Chinese*’, *Journal of the Chinese Language Teachers Association* XXIV.1: 43-61.

- Chao, Y.-R. (1930), 'A System of Tone Letters', *Le Maitre Phonétique* 30: 24-27. (Reprinted in regular English orthography in *Fangyan* (1980) 2: 81-83.)
- _____ (1933), 'Tone and Intonation in Chinese', *Bulletin of the Institute of History and Philology* 4: 121-34.
- _____ (1948), *Mandarin Primer* (Cambridge, MA: Harvard University Press).
- _____ (1968), *A Grammar of Spoken Chinese* (Berkeley: University of California Press).
- Chen, M.-Y. (2000), *Tone Sandhi: Patterns Across Chinese Dialects* (Cambridge, UK: Cambridge University Press).
- Cheng, R.-L. (1985), 'A Comparison of Taiwanese, Taiwan Mandarin, and Peking Mandarin', *Language* 61.2: 352-77.
- Cheng, C.-C. (1973), *A Synchronic Phonology of Mandarin Chinese* (The Hague: Mouton).
- Chou, F.-C., Tseng, C.-Y., and Lee, L.-S. (1996), 'Automatic Generation of Prosodic Structure for High Quality Mandarin Speech Synthesis' *Proceedings of the Fourth International Conference on Spoken Language Processing* (Philadelphia, Pennsylvania), 1624-27.
- _____ (1998), 'Automatic Segmental and Prosodic Labelling of Mandarin Speech Database', *Proceedings of the Fifth International Conference on Spoken Language Processing* (Sydney, Australia).
- Duanmu, S. (1990), 'A Formal Study of Syllable, Tone, Stress and Domain in Chinese Languages', Ph.D. dissertation, MIT, Cambridge, Mass. (Distributed by MIT Working Papers in Linguistics).
- _____ (1997), 'Recursive Constraint Evaluation in Optimality Theory: Evidence from Cyclic Compounds in Shanghai', *Natural Language and Linguistics Theory* 15: 465-507.

- Feifel, K.-E. (1994), *Language Attitudes in Taiwan: A Social Evaluation of Language in Social Change* (Taipei: The Crane Publishing Co., Ltd.).
- Fon, J. (2000), 'Production and Perception of Two Dipping Tones (T2 and T3) in Taiwan Mandarin', Manuscript, Department of Linguistics, The Ohio State University.
- _____ and Chiang, W.-Y. (1999), 'What does Chao have to say about tones? — a case study of Taiwan Mandarin', *Journal of Chinese Linguistics* 27.1: 15-37.
- Grice, M., D'Imperio, M., Savino M., and Avesani, C. (this volume), '**Towards a Strategy for Labelling Varieties of Italian**'.
- Grimes, B. F. (ed.) (1996), *Ethnologue: Languages of the World*, 13th edition (Dallas, Texas: Summer Institute of Linguistics) (The statistics on Mandarin speakers are from the online version <<http://www.sil.org/ethnologue/>>).
- Hartman, L. M. (1944), 'The Segmental Phonemes of the Peiping Dialect', *Language* 20: 28-42.
- Hsiao, Y.-C. E. (1991), *Syntax, Rhythm and Tone: A Triangular Relationship* (Taipei: The Crane Publishing Co.).
- Hu, M.-Y. (1987), 'Putonghua he Beijinghua' [Putonghua and Beijing Dialect], In *Beijinghua Chu Tan* [A Preliminary Study of the Beijing Dialect], by Hu, M.-Y. (Beijing: Commercial Press) (Reprinted in 1991 in *Yuyanxue Lunwenji* [English title: Selected Writings in Linguistics], by Hu, M.-Y., Beijing: Zhongguo Renmin Daxue Chubanshe, 167-87).
- Huang, T. (1999), 'A First Look at Rugao Chinese', paper presented at the Colloquiumfest, Department of Linguistics, The Ohio State University.
- Hung, T. T.-N. (1989), 'Syntactic and Semantic Aspects of Chinese Tone Sandhi', Ph.D. dissertation, University of California, San Diego (Reproduced by the Indiana University Linguistics Club, Bloomington, Indiana).

- Jiangsu Sheng he Shanghai Shi Fangyan Diaocha Zhidao Xiaozu (1960), *Jiangsu Sheng he Shanghai Shi Fangyan Gaikuang* [Survey on Dialects in Shanghai City and Jiangsu Province] (Nanjing: Jiangsu Renmin Chubanshe).
- Jin, S.-D. (1985), 'Shanghai Morphotonemics: A Preliminary Study of Tone Sandhi Behavior Across Word Boundaries', M.A. thesis, University of Pittsburgh, Pittsburgh (Reproduced by Indiana University Linguistics Club, Bloomington, Indiana).
- _____ (1996), 'An Acoustic Study of Sentence Stress in Mandarin Chinese', Ph.D. dissertation, The Ohio State University, Ohio.
- Kratochvil, P. (1968), *The Chinese Language Today: Features of an Emerging Standard* (London: Hutchinson University Library).
- _____ (1987), 'The Case of the Third Tone', In the Chinese Language Society of Hong Kong (eds.), *Wang Li Memorial Volumes, English Volume* (Hong Kong: Joint Publishing Co., 253-277).
- _____ (1998), 'Intonation in Beijing Chinese', In Daniel Hirst and Albert Di Cristo (eds.) *Intonation Systems: A Survey of Twenty Languages* (New York: Cambridge University Press, 417-431).
- Lee, O. J. (2000), 'The Pragmatics and Intonation of Ma-Particle Questions in Mandarin', M.A. thesis, The Ohio State University, Ohio.
- Li, A.-J., Zu, Y.-Q., and Li, Z.-Q. (1999), 'A National Database Design and Prosodic Labelling for Speech Synthesis', Paper presented at the 1999 Oriental COCODA Workshop on East Asian Language Resources and Evaluation (Taipei, Taiwan).
- Li, R. (1985), 'Guanhua Fangyan de Fenqu' [The Grouping of the Mandarin Dialects], *Fangyan* 1: 2-5.

- _____ (1989a), 'Zhongguode Yuyan he Fangyan' [Languages and Dialects in China], *Fangyan* 3: 161-67.
- _____ (1989b), 'Hanyu Fangyande Fenqu' [The Classification of the Chinese Dialects], *Fangyan* 4: 241-59.
- Liao, R.-R. (1994), 'Pitch Contour Formation in Mandarin Chinese: A Study of Tone and Intonation', Ph.D. dissertation, The Ohio State University, Ohio.
- Liberman, M. and Pierrehumbert, J. (1984), 'Intonational Invariance under Changes in Pitch Range and Length', in Aronoff, M. and Oerhle, R. (eds.), *Language and Sound Structure* (Cambridge, Mass.: MIT Press, 157-233).
- Peng, S.-H. (1996), 'Phonetic Implementation and Perception of Place Coarticulation and Tone Sandhi', Ph.D. dissertation, The Ohio State University, Ohio.
- Qian, Z.-Y. (1982), *Yantai Fangyan Baogao* [Report on the Yantai Dialect] (Ji'nan: Qilu Shushe).
- Qingdaoshi Shizhi Bangongshi (1997), *Qingdao Shizhi: Fangyan Zhi*. [Qingdao City Records: Dialect Records] (Beijing: Xinhua Chubanshe).
- Rugao Xianzhi Editing Committee (1994), *Fangyan* [Dialect], in Li, S.-Q., Cai, C.-M., and Shao, Y.-L. (eds.), *Rugao Xianzhi*. vol. 20 (Hongkong: New Asia Press).
- Shen, X.-N. (1989), 'Interplay of the Four Citation Tones and Intonation in Mandarin Chinese', *Journal of Chinese Linguistics* 17.1: 61-74.
- _____ (1990), *The Prosody of Mandarin Chinese* (Berkeley: University of California Press).
- Selkirk, E. and Shen, T. (1990), 'Prosodic Domains in Shanghai Chinese', in Inkelas S. and Zec D. (eds.), *The Phonology – Syntax Connection* (Chicago: The University of Chicago Press, 313-37).

- Shih, C.-L. (1986), 'The Prosodic Domain of Tone Sandhi in Chinese', Ph.D. dissertation, University of California, San Diego.
- _____ (1988), 'Tone and Intonation in Mandarin', *Working Papers of the Cornell Phonetics Laboratory* 3: 83-109.
- _____ (1997), 'Mandarin Third Tone Sandhi and Prosodic Structure', in Wang, J.-L. and Smith N. (eds.), *Studies in Chinese Phonology* (Berlin and New York: Mouton de Gruyter, 81-123).
- Stimson, H. M. (1966), 'A Tabu Word in the Peking Dialect', *Language* 42.2: 285-94.
- Tai, J. H.-Y. (1976), *Lexical Changes in Modern Standard Chinese in the People's Republic of China Since 1949* (Washington, D.C.: Office of Research, United States Information Agency).
- _____ (1977), *Syntactic and Stylistic Changes in Modern Standard Chinese in the People's Republic of China Since 1949* (Washington, D.C.: Office of Research, United States Information Agency).
- _____ (1978), *Phonological Changes in Modern Standard Chinese in the People's Republic of China Since 1949* (Washington, D.C.: Office of Research, United States Information Agency).
- Ting, P.-H. (1966), 'Rugao Fangyan de Yinyun' [Phonology of the Rugao Dialect], *Bulletin of the Institute of History and Philology* 36: 573-633.
- Tseng, C.-Y. and Chou, F.-C. (1999a), 'Machine Readable Phonetic Transcription System for Chinese Dialects Spoken in Taiwan', *J. Acoust. Soc. Jpn. (E)* 20.3: 215-23.
- _____ (1999b), 'A Prosodic Labeling System for Mandarin Speech Database', *Proceedings of the XIVth International Congress of the Phonetic Sciences*, vol. 3, 2379-82.
- Venditti, J. (this volume), '[The J_ToBI Model of Japanese Intonation](#)'.

- Wang, W. S.-Y. and Li. K.-P. (1967), 'Tone 3 in Pekingese', *Journal of Speech and Hearing Research* 10: 629-36.
- Wong, P. W.-Y., Chan, M. K.-M., and Beckman, M. E. (this volume), '**An Autosegmental-Metrical Analysis and Prosodic Annotation Conventions for Cantonese**'.
- Wright, M. (1983), 'A Metrical Approach to Tone Sandhi in Chinese Dialects', Ph.D. dissertation, University of Massachusetts, Mass.
- Xu, Y. (1999), 'Effects of Tone and Focus on the Formation and Alignment of F0 Contours', *Journal of Phonetics* 27.1: 55-105.
- Yin, Z.-Y. (1982), 'Guanyu Putonghua Shuangyin Changyongci Qingzhongyinde Chubu Kaocha' [A Preliminary Study of Accents and Atonics in Disyllabic Words in Common Use], *Zhongguo Yuwen* 3: 168-73.
- Yip, M. (1980), 'The Tonal Phonology of Chinese', Ph.D. dissertation, MIT, Cambridge, Mass.
- Yuan, J.-H. (1960), *Hanyu Fangyan Gaiyao* [Outline of Chinese Dialects] (Beijing: Wenzhi Gaige Chubanshe) (Second edition published in 1989).
- Zhang, Z.-S. (1988), 'Tone and Tone Sandhi in Chinese', Ph.D. dissertation, The Ohio State University, Ohio.

CHAPTER TABLES:

Table 11.1. Contrastive tones in standard Putonghua Mandarin

Tone	Description	Transcriptions		Example	Pinyin
Tone 1	high level	55	H	ba55 ‘a scar’	bā
Tone 2	high rising	35	LH	ba35 ‘to uproot’	bá
Tone 3	low falling	21	L	ba21 ‘a target’	bǎ
Tone 4	high falling	51	HL	ba51 ‘a dam’	bà
Neutral tone				ba ‘(particle)’	ba

Table 11.2. Contrastive tones in Rugaohua Mandarin

Tone	Description	Transcriptions		Example
Tone 1	falling	41	HL	ba ⁴¹ ‘a scar’
Tone 2	rising	35	MH	pa ³⁵ ‘to crawl’
Tone 3	low (falling-rising)	323	MLM	ba ³²³ ‘a target’
Tone 4	high level	55	H	ba ⁵⁵ ‘a dam’
Yin-ru	checked level	5	H	baʔ ⁵ ‘to shell’
Yang-ru	checked rising	35	MH	paʔ ³⁵ ‘thin’
Neutral tone				ba ‘(particle)’

Table 11.3. The eight tiers

Tier	Description
1. Words	syllable-by-syllable transcription in Chinese characters
2. Romanisation	syllable-by-syllable transcription in modified Pinyin romanisation using ASCII (hence, without tone diacritics)
3. Syllables*	phonological syllables that do not correspond in a one-to-one relationship with orthographic syllables (e.g., for contractions as such as <i>tām</i> for the disyllabic pronoun, <i>tāmen</i> ‘they’).
4. Stress	relative degree of stress marked on each syllable
5. Sandhi	marking of tone sandhi (e.g., 35 for the third tone sandhi form of the first syllable in /yu21san21/), allotones (e.g., 214 for Tone 3 in an utterance final position), morpheme-specific tone sandhi (e.g., <i>bu35</i> ‘not’ in <i>bú yào</i> ‘don’t want’, from <i>bù</i> + <i>yào</i>)
6. Tones	marking of boundary tones and pitch range effects (e.g., %compressed at the onset of pitch range reduction)
7. Break Indices	hierarchy of disjunctures to represent prosodic phrasing
8. Code	identification of the variety of Mandarin and marking of points of code-switching

* In the Emu implementation of the M_ToBI standard, the stress tier and sandhi tier are simply two extra labeling fields for the syllables level, whereas in the xwaves/xlabel implementation, the relationship among these tiers must be insured by an independent grammar checker.

Table 11.4. Stress levels in the stress tier

S3	syllable with fully-realised lexical tone
S2	syllable with substantial tone reduction (e.g., undershooting of tonal target with duration reduction)
S1	syllable that has lost its lexical tonal specification (e.g., in a weakly-stressed position)
S0	syllable with lexical neutral tone (i.e., such a syllable is inherently unstressed)

Table 11.5. Tags for boundary tones and backdrop pitch range effects in the tones tier

H%	high boundary tone (at the end of an utterance)
L%	low boundary tone (at the end of an utterance)
%reset	beginning of a new pitch downtrend or pitch reset
%q-raise	beginning of a raised pitch range (e.g., in echo questions)
%e-prom	beginning of local expansion of pitch range due to emphatic prominence
%compressed	beginning of reduction of pitch range of syllables following the expansion of pitch range under %e-prom

Table 11.6. Break indices and corresponding events on other tiers

B0	reduced syllable boundary	i.e. contraction; requires S0 or S1 on left or right
B1	normal syllable boundary	the “default” case within a polysyllabic word
B2	minor phrase boundary	must be followed by at least S2
B3	major phrase boundary	
B4	breath group boundary	H%, L%, %reset, and %q-raise on the tones tier
B5	prosodic group boundary	

CHAPTER FIGURES:

Type setters: We have made all figures 4.5 inches wide, so that they need not be reduced.

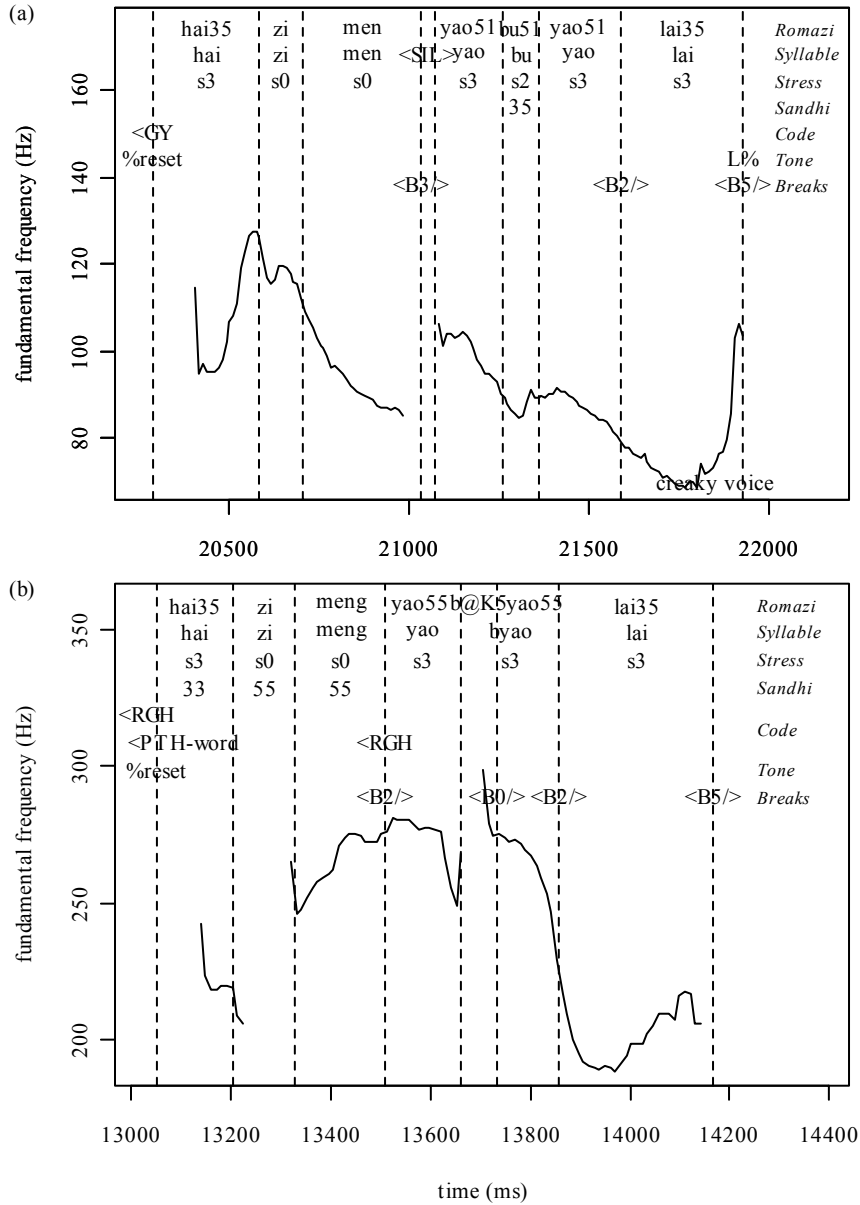


Figure 11.1. (a) Guoyu utterance of *Hái zǐ mēn yào bù yào lái?* ‘Do the children want to come?’
 (b) Rugaohua utterance of same sentence as in Figure 11.1a.

Typesetters: The two panels of this figure should just fit on one page along with the figure caption.

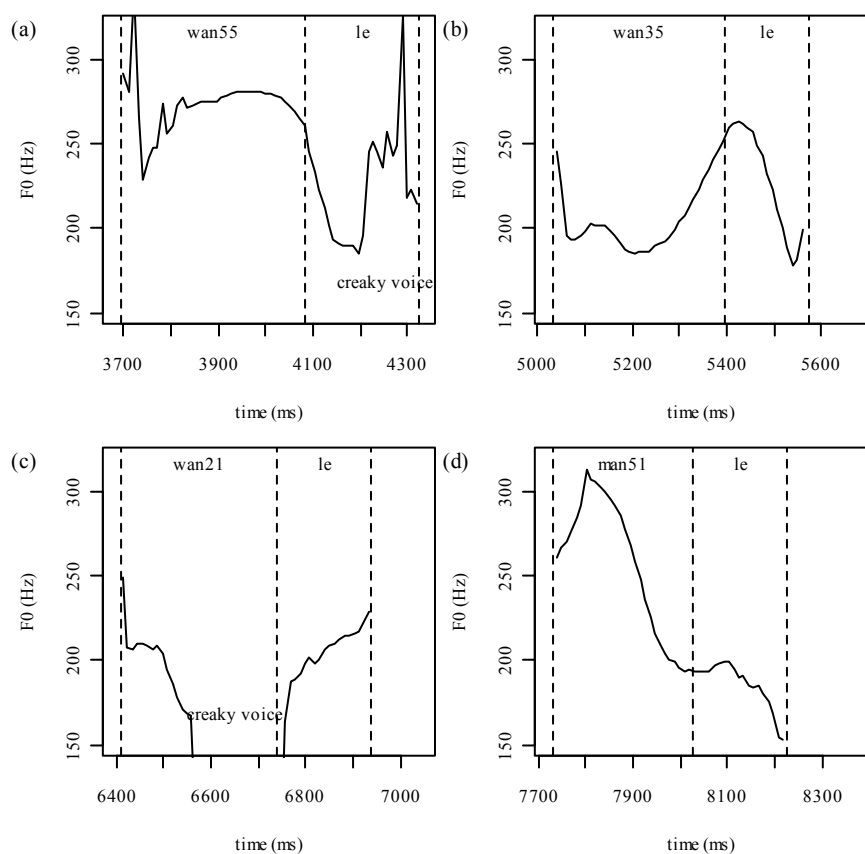


Figure 11.2. Putongua utterances of disyllabic predicates illustrating the realization of the neutral tone on the aspect particle *le* after each of the four full tones: (a) *wān le* ‘to have (been) bent.’ [neutral tone after Tone 1], (b) *wán le* ‘to have finished.’ [neutral tone after Tone 2], (c) *wǎn le* ‘to have been late.’ [neutral tone after Tone 3], and (d) *màn le* ‘to be too slow.’ [neutral tone after Tone 4].

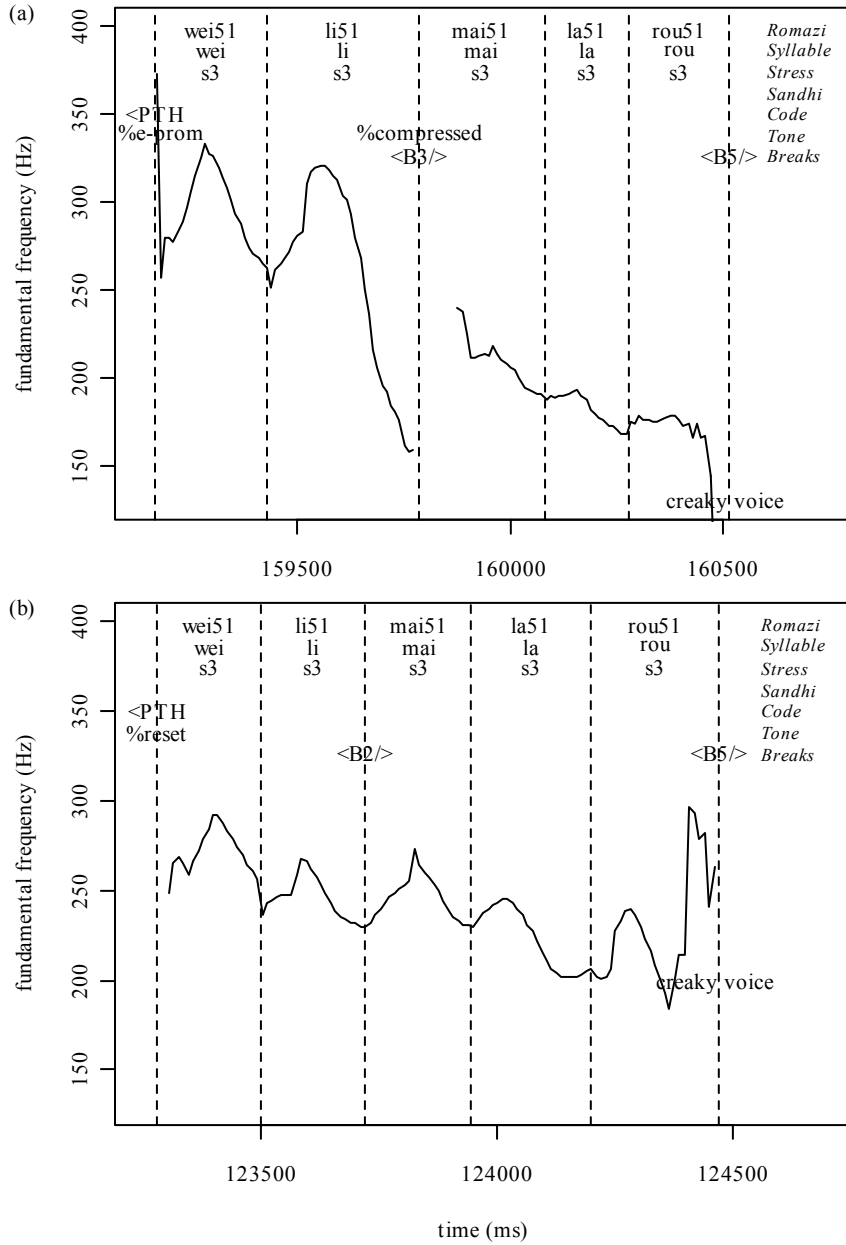


Figure 11.3. (a) Putonghua utterance of sentence *Wèi Lì mài là ròu*. ‘Wei Li sells (Chinese) bacon.’ as a statement with narrow focus on *Wèi Lì*.
 (b) The same sentence uttered as an “out of the blue” statement with broad focus.

Typesetters: Figure 11.3, first page -- this figure will need to continue onto a second page. We have put an x-axis label only on panel (b), to have only one on each of the 2 pages. Therefore, layout should be 2 panels and then 1, as here, rather than 1 and 2.

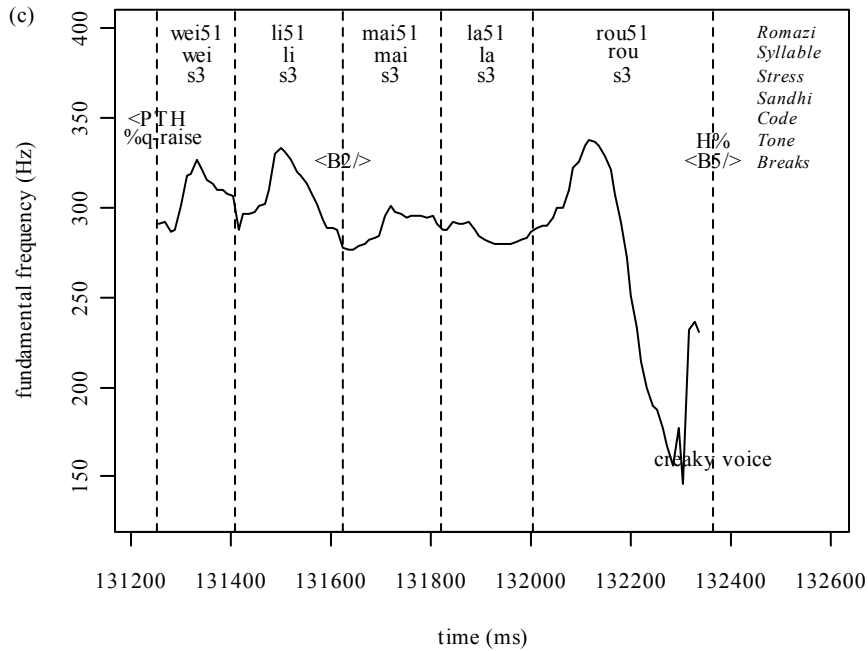


Figure 11.3. (c) The same speaker producing the sentence in Figure 11.3a as an echo question: ‘Wei Li sells (Chinese) bacon!?’

Typesetters: Figure 3, second page -- Perhaps put the larger figure caption here, and a small figure caption on the first page where panels (a) and (b) of the graph are printed?

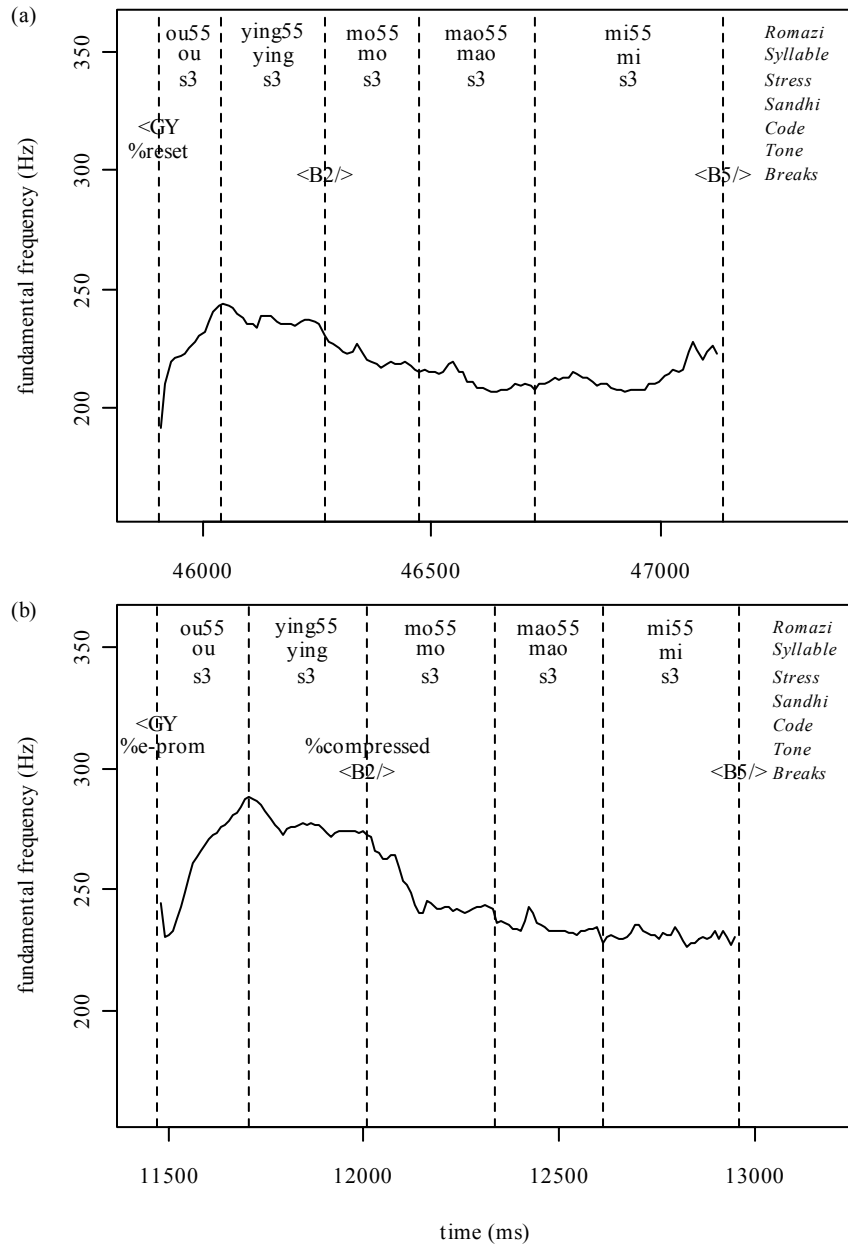


Figure 11.4. (a) Guoyu broad-focus utterance of $\bar{O}u\ y\bar{i}ng\ m\bar{o}\ m\bar{a}o\ m\bar{i}$. ‘Ouying strokes kitty.’
 (b) The same speaker’s utterance of the sentence with narrow focus on $\bar{O}u\ y\bar{i}ng$.

Typesetters: Figure 11.4, first page -- Figure will have to be spread over two pages. We have put x-axis labels only on panels (b) and (d), so that the layout can be two panels per page as shown.

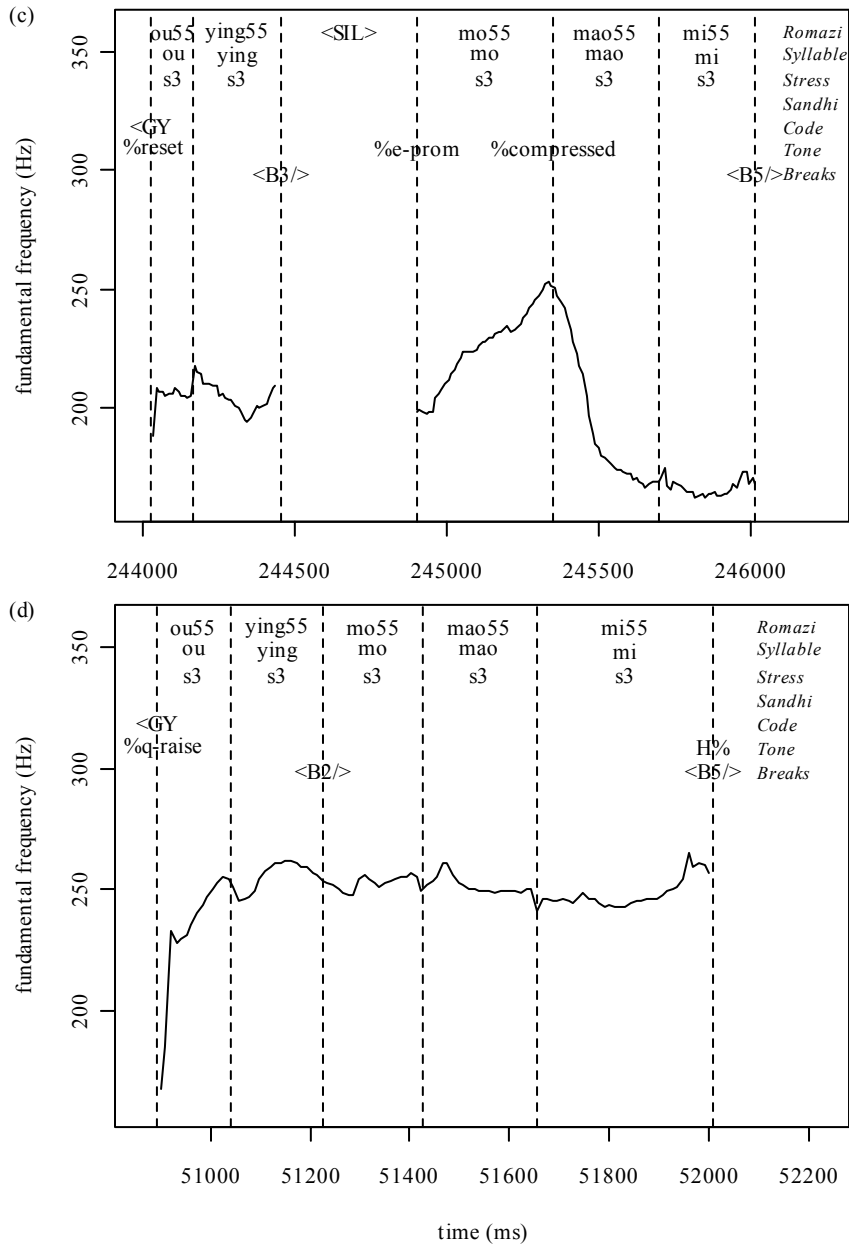


Figure 11.4. (c) The same speaker's utterance of the sentence with narrow focus on *mō*, and pause before it.
 (d) The same speaker producing the sentence in Figure 11.4a as an echo question.

Typesetters: Figure 11.4, second page. See Figure 11.4, first page, for suggested layout.

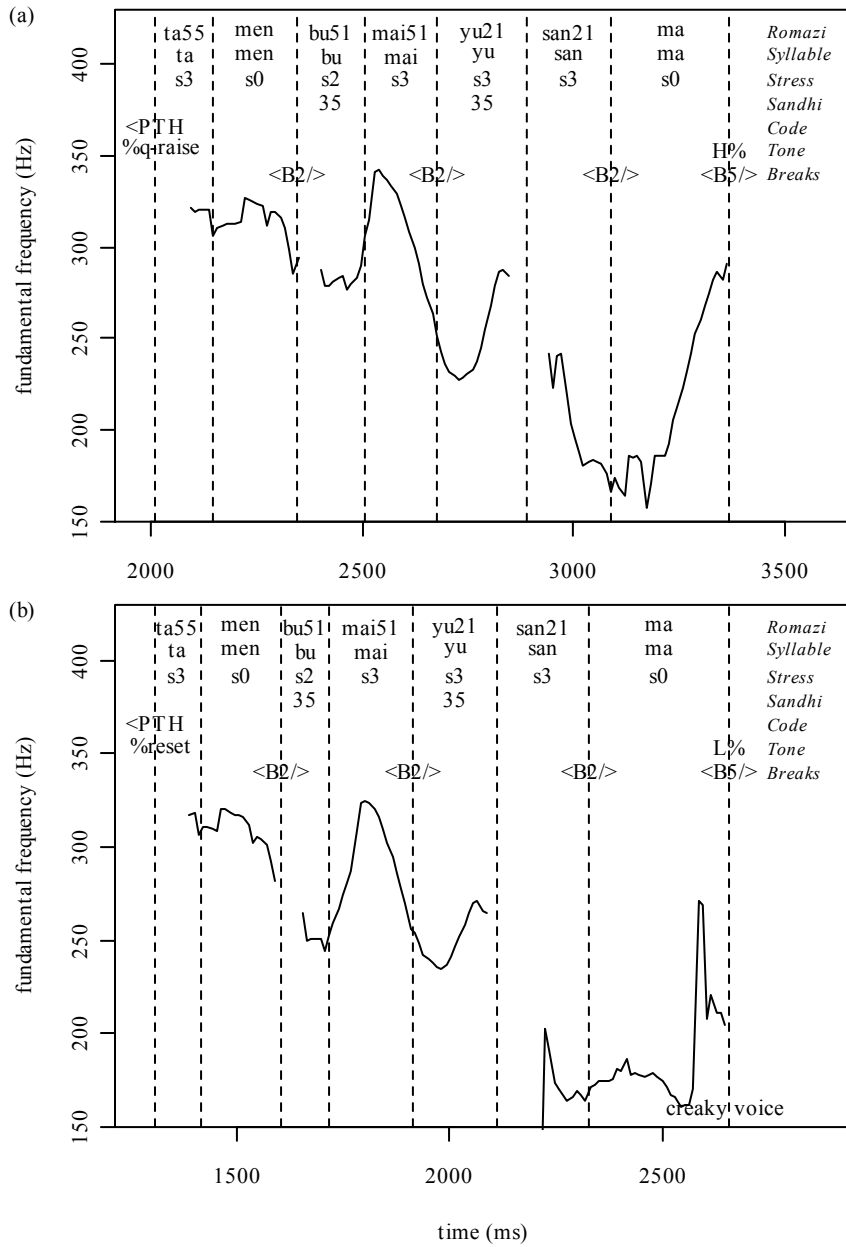


Figure 11.5. (a) Putonghua utterance of *Tā men bù mài yǔ sǎn ma?* 'They don't sell umbrellas?' produced with H%.
 (b) The same utterance as Figure 5a produced with L%.

Typesetters: See Figure 11.1 for suggested layout.

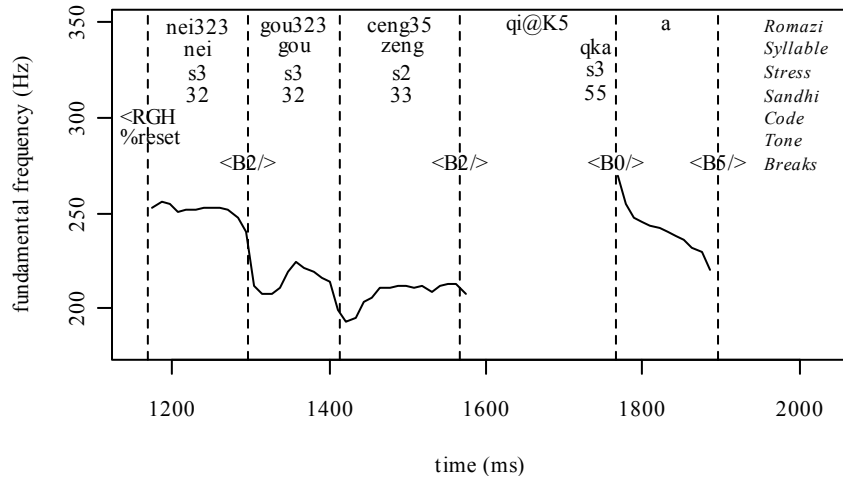


Figure 11.6. Rugao utterance of the sentence *Nei323 gou323 ceng35 qi@K5 a*. ‘Have you eaten?’ with tone spreading from *gou323* to *ceng35*, and from *qi@K5* to *a*.

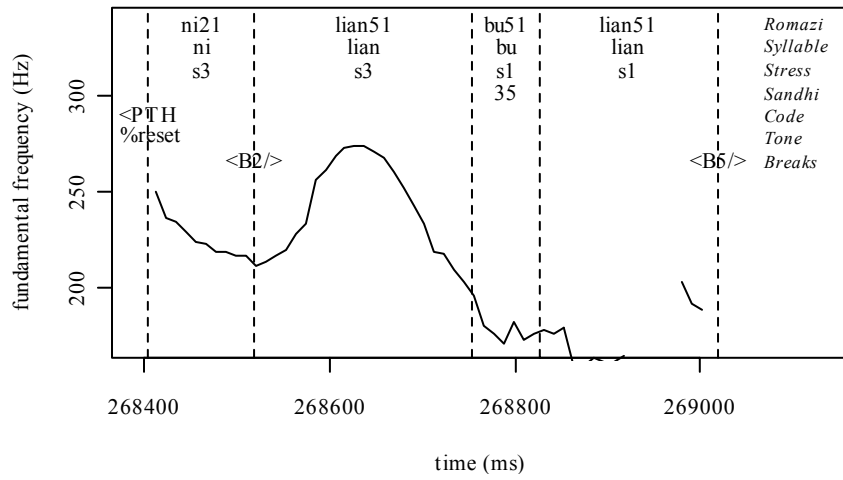


Figure 11.7. Putonghua utterance of *Nǐ liàn bù liàn?* ‘Will you be practising?’

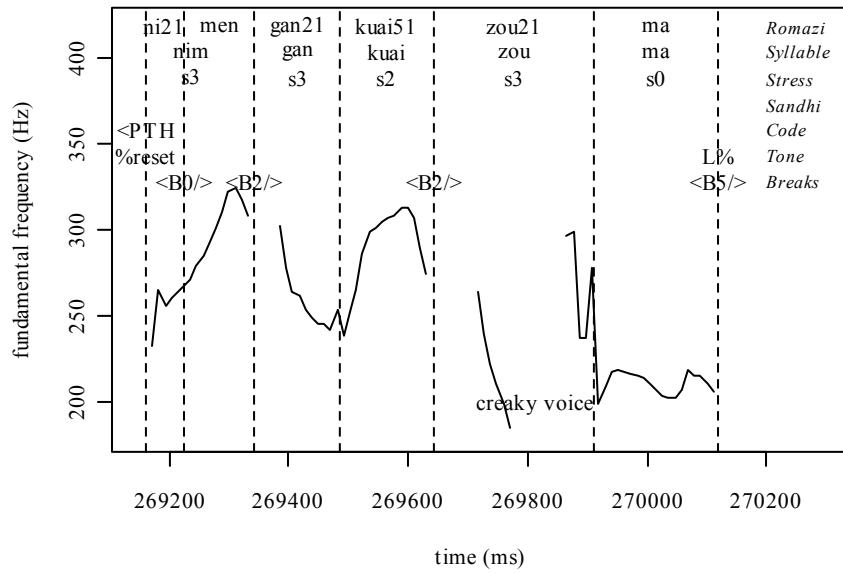


Figure 11.8. Putonghua utterance of the sentence *Nǐ men gǎn kuài zǒu ma!* ‘You (pl.) should leave right away!’

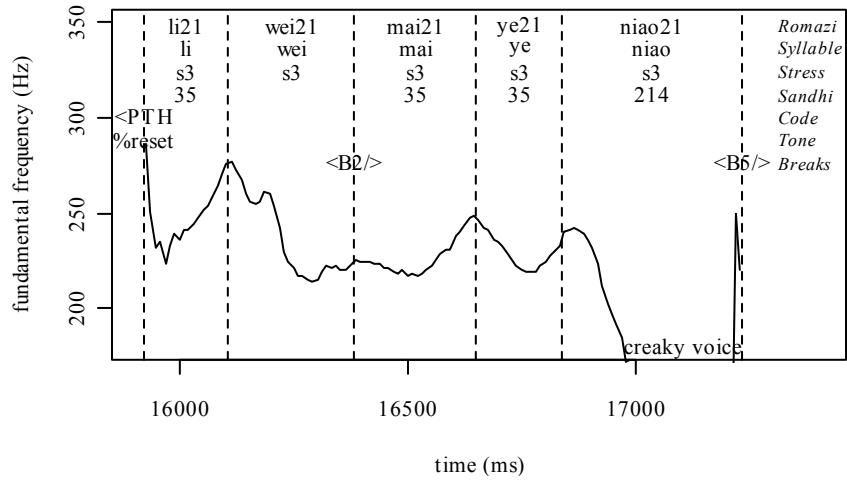


Figure 11.9. Putonghua utterance of *Lǐ Wěi mǎi yě niǎo*. ‘Li Wei buys a wild bird.’

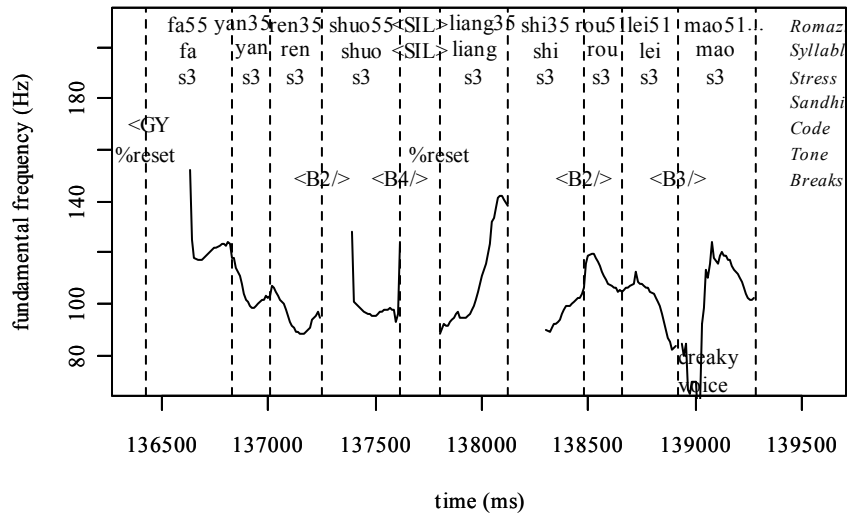


Figure 11.10. Fragment of Guoyu utterance *Fā yán rén shuō: Liáng shí ròu lèi mào yì yáng háng jiāng yú míng rì zhèng shì kāi yè.* ‘The spokesperson said: “Farmers and Butchers Trading Company will open for business tomorrow.”’

Abstract

This article describes the initial stages of development of a Pan-Mandarin ToBI system. We review the salient prosodic characteristics of Mandarin, such as lexical tones, tone sandhi, tonal neutralisation, stress patterns, pitch range effects, and prosodic groupings above the syllable level. Particular attention is paid to the range of variability within a common structural core, in addition to points of reference to other non-Mandarin varieties of Chinese and to other languages that have been described within the ToBI framework. We then propose a codification of conventions for marking prosodic structure, and an inventory of tones, as well as other structure-marking elements in two standard varieties (i.e., Putonghua of Mainland China and Guoyu of Taiwan) and one regional variety of the language (i.e., Rugaohua, a Jianghuai Mandarin variety). We aim for the Pan-Mandarin ToBI system to accommodate our expanding knowledge of more regional varieties as well as of the standard varieties. Also built into the system is the capability to accommodate interactions, such as code-switching events, between different varieties of Mandarin and perhaps between Mandarin and other varieties of Chinese (and other languages) in different social contexts.

Keywords:

Mandarin
Chinese
tones
stress
intonation
tone sandhi
Rugaohua
Guoyu
Putonghua
M_ToBI

Index:

Putonghua
Guoyu
Rugaohua
tone sandhi
neutral tone
stress
lexical tones
narrow focus
broad focus
boundary tones
prosodic grouping
pitch range
tone neutralization
code-switching
third tone sandhi