CHAPTER 36

INTONATION IN CHINESE

YI XU

Intonation refers to fundamental frequency (F_0) patterns in speech that convey information beyond lexical meanings. In a tone language, F_0 is already used to represent lexical tones. An obvious question is therefore how it is possible for it to also carry intonation. A widespread idea is that the coexistence of tone and intonation in a tone language is possible because tones are like small ripples riding on large intonation waves (Chao 1968). This chapter summarizes the scientific findings that, though consistent with this view, show specific mechanisms that make it possible for tone and intonation to coexist. Importantly, these findings have shown that an adequate understanding of intonation is achievable only if we take into account two complementary aspects of speech: articulatory mechanisms and communicative functions. The chapter first outlines the basic mechanisms of pitch production, which serves as the basis of explaining how intonational melody can be realized in parallel with tones. The intonational components are then introduced in terms of their functions with specific communicative meanings, which include focus, sentence modality, topic, turn taking, and boundary marking. The relation between these functions and the often-reported phenomena of declination, downstep, and rhythm are also discussed.

36.1 Tonal Mechanisms

F_0 is produced by the vibration of the vocal folds. The rate of this vibration is determined jointly by the tension of the vocal folds and subglottal pressure (Titze 1989). The former is mainly controlled by the intrinsic and extrinsic laryngeal muscles and the latter by the abdominal muscles and the diaphragm in conjunction with the laryngeal muscles (Zemlin 1988; Titze 1989). A critical outcome of this complex control is the time course and dynamics of the resulting F_0 movements. Two patterns have emerged from empirical research. First, it takes a significant amount of time for a speaker to make even the smallest pitch shift, and the time needed increases quasi-linearly with the size of

pitch change, as represented by the following equations based on data from Xu and Sun (2002):

$$t = 89.6 + 8.7d \quad \text{(raising)} \qquad [1]$$

$$t = 100.4 + 5.8d \quad \text{(lowering)} \qquad [2]$$

where t is the amount of time (millisecond [ms]) it takes for an average speaker to complete an F_0 movement and d the size of the F_0 movement in semitones (st = 12 * log2 (F_0 / F_0reference)). According to [1] and [2], it takes about 124 ms to either raise or lower pitch by 4 st. In continuous speech, the difference between two adjacent tones is often close to or greater than 4 st (Xu 1997, 1999). On the other hand, the duration of a syllable is very short when compared to the minimum time needed to make the tonal transitions. In Mandarin, for example, the average syllable duration is about 180 to 215 ms (Duanmu 1994; Xu 1999). This means that to make a shift between two adjacent tones, a time period greater than half of a syllable has to be used just for making the transition even if the maximum speed of pitch change is achieved.

The second pattern emerging from empirical research is that the transition toward each tonal target occurs *within* the tone-bearing syllable itself (Xu 1997) rather than in a temporal interval *between* adjacent tones (Ladd 2008). Combined with the relatively slow F_0 movements, the F_0 of an entire syllable constitutes a continuous transition toward an ideal pitch pattern of the tone, as shown in Figure 36.1.

These findings suggest that *syllable-synchronized sequential target approximation* is the basic mechanism of tone articulation, as summarized by the Target Approximation (TA) model (Xu and Wang 2001) illustrated in Figure 36.2. Under the TA model, pitch targets serve as *melodic segments* whose presence and implementation are mandatory (i.e., required whether or not the language is tonal). On the other hand, various aspects of the target approximation process can be adjusted, and the adjustments can determine additional details of the surface F_0 contours. At least four aspects of the target approximation process, including the targets themselves, can be identified: (i) pitch target (specified in terms of slope and height), (ii) strength (rate of target approximation), (iii) pitch range (height and span), and (iv) syllable duration. The impact of modifying each of the four parameters can be seen in Figure 36.2.

36.1.1 Tonal Pitch Range versus Total Pitch Range

A relevant question about the possibility of using pitch range to represent information beyond lexical tones is whether there is actually sufficient extra pitch range available for this purpose. The question is important given that when describing lexical tones, it is often assumed that tones take up "the range of a speaker's voice" (Chao 1968:25). In Chao's 5-point scale system, for example, Tone 1 (high-level) in Mandarin is at the

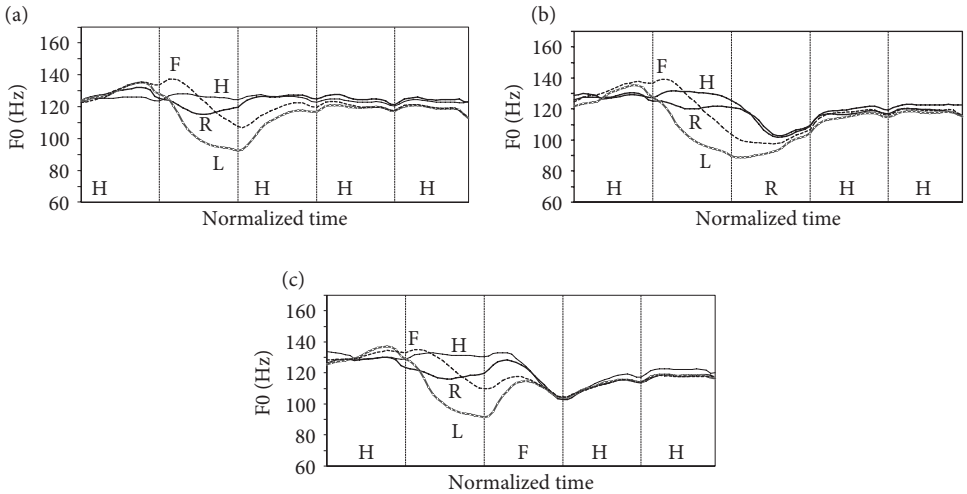

FIGURE 36.1 Mean time-normalized F_0 contours of 猫咪/迷/米/蜜 摸/拿/卖猫咪 [Kitty/Cat-fan/Cat rice/Cat honey stroke/pick up/sell kitty] produced by four males speakers of Beijing Mandarin (Xu 1999). In all plots, vertical lines indicate syllable boundaries. Each contour is an average of forty tokens said by four male speakers of Beijing Mandarin (five repetitions by each). Adapted from Xu (1999).

top of the pitch range and Tone 3 (low-dipping) is at the bottom of the pitch range. Experimental data have shown that at any particular sentence position, the pitch range across the four Mandarin tones spans no more than one octave (1 octave = 12 semitones; Xu 1999). On the other hand, according to Fairbanks (1959), a speaker's conversational pitch range can span *two* octaves. According to Honorof and Whalen (2005), an average English-speaking male or female can produce a nonfalsetto voice covering a pitch span of as much as *three* octaves. Thus in a tone language like Mandarin, well over a full octave of the speakers' total pitch range, mostly in the upper region, is unused by lexical tones, hence available for other purposes. This suggests that there is actually plenty of room for F_0 to carry information beyond lexical tones.

36.2 Intonational Functions

Articulatory mechanisms constitute only one of the two complementary aspects of intonation. The other aspect consists of communicative functions, that is, the information that is actually conveyed to the listener. Like lexical tone, which serves to ~~separate~~ words or syntactic categories, intonation consists of individual functions that each serve to mark a specific meaningful contrast. It is therefore important to identify them individually and to determine, in each case, the specific encoding schemes and the interaction between different functions. Based on the most up-to-date research, the intonational

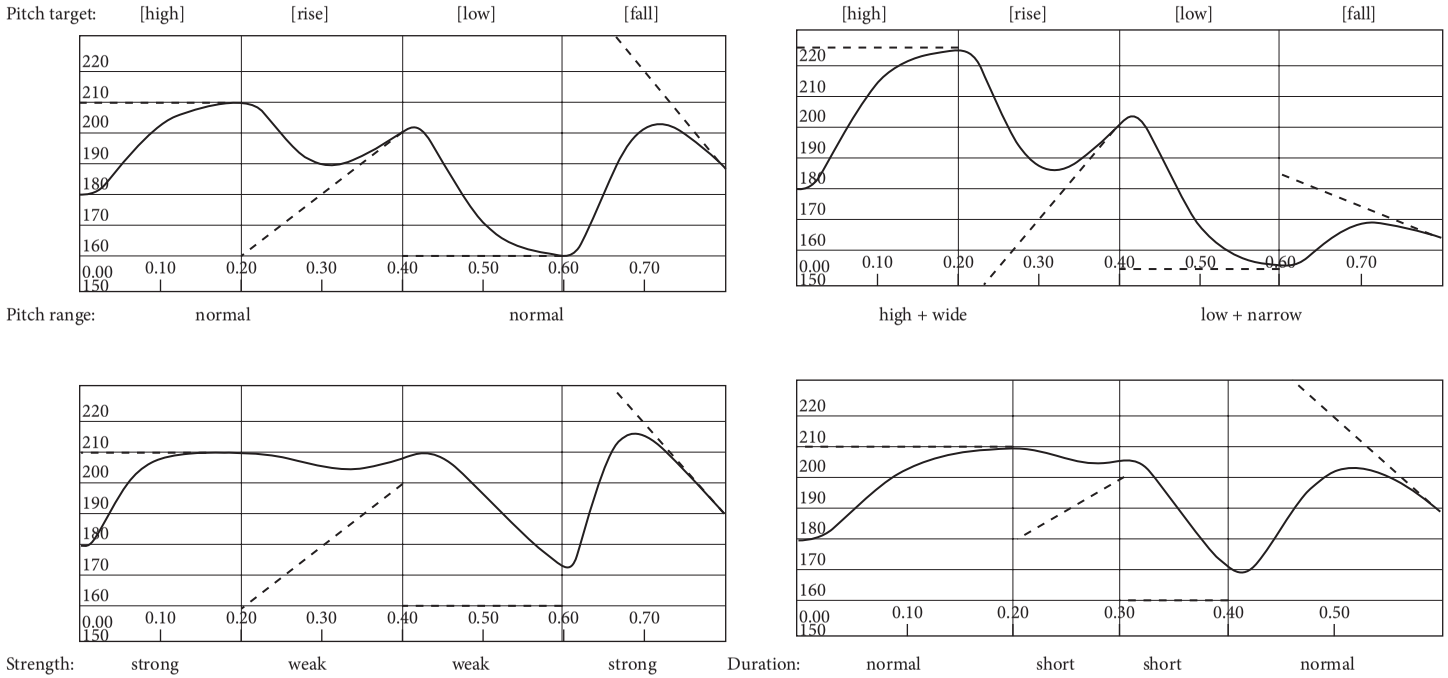

FIGURE 36.2 The TA model of tone production (Xu and Wang 2001), simulated by the qTA model (Prom-on et al. 2009). In all plots, vertical lines indicate syllable boundaries. The straight dashed lines represent pitch targets associated with lexical tones: static [high], dynamic [rise], static [low], and dynamic [fall]. The solid curves represent surface F_0 contours resulting from continuous articulatory approximation of successive pitch targets. The top left plot illustrates the case of normal lexical tone. The other three plots illustrate the effects of modifying *pitch range, strength,* and *syllable duration*, respectively. All the plots were generated by an interactive Java program accessible at http://www.phon.ucl.ac.uk/home/yi/qTA/.

functions recognized for Chinese languages include focus, modality, topic, and boundary marking.

36.2.1 Focus

When the sentence 他[he]昨天[yesterday]走了[left] [He left yesterday] is said as a response to the question 谁[who]昨天走了? [Who left yesterday?], the word 他 is naturally emphasized. When it is in response to 他什么时候[when]走了? [When did he leave?] or他昨天怎么了[what happened]? [What happened to him yesterday?], the focus would instead be on 昨天 or 走了. Such an emphasis, which highlights a particular component against the rest of the utterance, is known as *prosodic focus, narrow focus*, or simply *focus*. Prosodic focus has been extensively studied, and the emerging picture is that it is marked by a multitude of cues. In general, a focused component exhibits expanded pitch range, longer duration, greater intensity, and possibly increased high-frequency spectral energy (Cooper et al. 1985; Sluijter and van Heuven 1996; Xu 1999; de Jong 2004). Here pitch range expansion refers to the widening of the entire tonal pitch range, with the High tone becoming higher and the Low tone becoming lower. Furthermore, it is not just the focused component alone but also other parts of the sentence that are modified by the focal emphasis. In particular, the pitch range and intensity of the post-focus region are compressed (narrowed and lowered), hence, *post-focus compression* (PFC). But little is systematically changed in prefocus words. The asymmetrical pitch range change around focus can be seen in Figure 36.3b, where the High tone becomes higher and the Low tone becomes lower under focus, and all the postfocus tones become lower and reduced, as compared to the neutral focus F_0 contours in Figure 36.3a.

The identification of asymmetrical pitch range control by focus in Mandarin and English, in which the most notable feature is PFC, has led to a significant new development in intonation research. That is, PFC turns out to be nonuniversal, and it is unevenly distributed among the world languages (Xu 2011). Interestingly, this is even true of Chinese languages. For example, unlike Beijing Mandarin, Southern

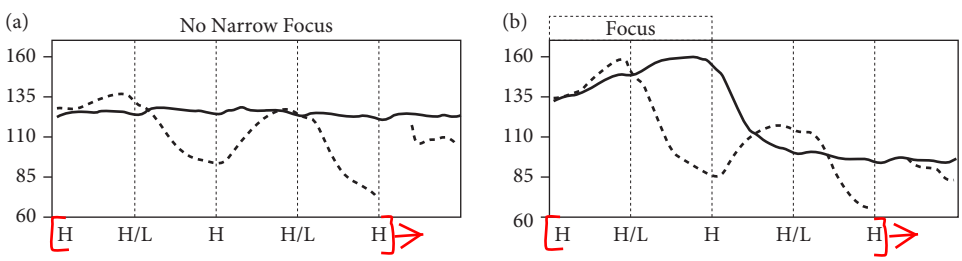

FIGURE 36.3 Mean time-normalized F_0 contours of 猫咪摸猫咪 [Kitty strokes kitty] (solid line) and猫米摸马刀 [Kitty strokes saber] (dotted line) with either no narrow focus (a) or focus on the first word (b). Each contour is an average of forty tokens said by four male speakers of Beijing Mandarin (five repetitions by each). Adapted from Xu (1999).

Min, Cantonese, and even Taiwan Mandarin (Mandarin spoken in Taiwan) all lack PFC (Xu et al. 2012). In addition, PFC is also absent in Yi, Deang, and Wa but present in Tibetan, Uygur, and Nanchang (Wang et al. 2011). More globally, there is initial evidence that PFC distribution correlates with language families. Among the PFC families are Indo-European, Uralic, and Altaic. The case of the Sino-Tibetan family is special, as PFC occurs in only some of the member languages (Beijing Mandarin, Tibetan, Nanchang) but is absent in others (Southern Min, Cantonese, Taiwan Mandarin). This may raise questions about the relation among these languages as well as their relation to other language families. A new area of research therefore seems to be emerging that may lead to rethinking about language typology, language change, and human evolution (Xu 2011).

36.2.2 Modality and Utterance Final Particles

Modality refers to whether an utterance is said as a statement or a question. This function is often marked morphosyntactically and in Chinese languages often by sentence-final particles (SFPs). In Mandarin, as in many other languages, a question is known to have a rising intonation toward the end of the utterance, whether or not a question particle is present. However, it is also known that in spontaneous speech even a syntactic question may not necessarily have a rising intonation and a syntactic statement may not have a falling intonation (Hirschberg 2000). Thus there are two separate issues about question intonation. One is *when exactly it occurs*, for which there is little consensus and therefore much more research is needed. The other is *what exactly happens* when it does occur, and this can be empirically studied. Interestingly, in an experimental recording session, a speaker would automatically use a rising intonation to read aloud a sentence with a question mark. This allows the use of systematic acoustic analysis to establish the prosodic details of question intonation. The general finding so far is that, at least in Mandarin, modality interacts extensively with focus to determine not only the utterance-final but also the entire F_0 contour of a sentence (Yuan 2004; Liu and Xu 2005). Figure 36.4 shows mean F_0 contours of a Mandarin question versus statement in sentences consisting of only the High tone. In Figure 36.4a, focus is either on the sentence-initial word or there is no narrow focus (hence, neutral focus). In Figure 36.4b, focus is either sentence-medial or sentence-final. It is apparent that the divergence between statement and question starts from the focused word rather than occurring only in the final word of the sentence or from the beginning of the sentence.

The F_0 patterns in Figure 36.4 also show that the pitch range adjustment by question intonation is a nonlinear function of time, with greater increase toward the end of the question. This nonlinearity is consistent with the general observation that the final raising is typically much more conspicuous than F_0 changes in earlier regions. But the interaction of modality with focus shows that the right edge of an utterance is not the only place where modality coding happens. This means that the notion of boundary tone (Pierrehumbert 1980), which implies that final F_0 rise can exist by itself, is not accurate,

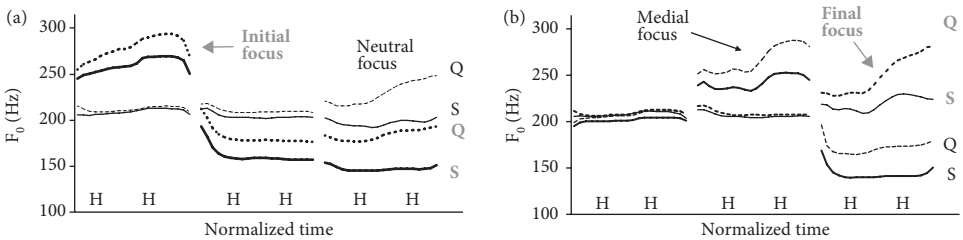

FIGURE 36.4 Mean F_0 contours of Mandarin sentence 张威 [Zhang Wei] 担心 [is concerned] 肖英 [Xiao Ying] 开车 [Driving] 发晕 [get dizzy] [Zhang Wei is concerned that Xiao Ying may get dizzy when driving] spoken as either a statement or a question. On the left, either focus is on the sentence initial word, or there is no narrow focus. On the right, focus is either sentence medial or sentence final. Data from Liu and Xu (2005).

because it is only part of the pitch range adjustment by modality, and additional adjustments are applied to other parts of the sentence, especially at the location of focus. Furthermore, the nonlinear adjustment of pitch range occurs not only in questions but also in statements, where F_0 is increasingly *lowered* toward the end of an utterance, resulting in the greatest drop in the final syllable. This final drop is observed as final lowering for English (Liberman and Pierrehumbert 1984).

The accelerating F_0 change found in Mandarin reflects the importance of sentence-final location for marking modality-related information. Beside question/statement contrast, other information can be encoded also at that location. This is reflected in the wide use of SFPs among the Chinese languages. One of the richest repertoires of SFPs can be found in Cantonese. A recent study has shown that Cantonese SFPs make use of a combination of lexical-tone-like F_0 contours and modality-relevant modifications such as question (Wu 2009). Also, some of the Cantonese SFPs exhibit an edge tone like component, which is sequentially attached to the end of the SFP by lengthening the rhyme without adding an extra syllable. Similar edge tones were noted in Mandarin by Chao (1968), who referred to them as particles. Only recently has there been some empirical research on the acoustic characteristics of this component (Mueller-Liu 2006; Li et al. 2011). More research is certainly needed.

36.2.3 New Topic (Topic Shift / Turn Taking)

When a sentence is the very first in a conversational turn or a read paragraph, its F_0 is raised, sometimes extensively, relative to other sentences (Lehiste 1975; Umeda 1982). The raising seems to affect the pitch range of an utterance nonlinearly, with a large F_0 raise near the beginning and a gradual drop afterward (Wang and Xu 2011). The gradual drop clearly separates topic from focus, which immediately lowers postfocus pitch range, as shown in Figure 36.3b. The communicative function of this initial F_0 raise is not yet fully clear. It has been described as marking a new paragraph (Lehiste 1975;

Tseng et al. 2005), introducing a new topic (Nakajima and Allen 1993; Wang and Xu 2011), or initiating a new conversational turn (Swerts 1997). Conceptually, topic seems to overlap with some other prosodic notions, particularly with focus and newness. The conceptual overlap with focus is the greatest in the case of contrastive topic, which is described as a topic that contains an alternative and marked by a rising accent in English (Büring 2003). So far, however, there have been no experimental data in support of a rising accent associated with contrastive topic in English. For Mandarin, a recent study has found no acoustic cue for the contrastiveness in a topic (Wang and Xu 2011). The same study has further found that after controlling for focus and topic, newness ~~also~~ does not have intonational correlates except for a slightly lengthened duration (Wang and Xu 2011).

36.2.4 Boundary Marking

The importance of boundary marking is exemplified by the ancient punctuation story of 下雨天留客天天留我不留, which can be interpreted as either 下雨天[A rainy day], 留客天[is a day for keeping visitors], 天留我不留[The weather keeps the visitor but I don't], or 下雨天[A rainy day], 留客天[is a day for keeping visitors], 天留我不[Does the weather keep me]? 留[Yes]. In spoken language, such punctuation can be achieved through boundary marking by pausing. Pausing, however, does not always involve silence. Rather, domain-final lengthening can serve as a boundary marker. For English, it is shown that syllable duration alone can mark as many as seven grades of boundary strength (Wagner 2005). For Mandarin, domain-final lengthening is also consistently found (Tseng et al. 2005; Yuan et al. 2006; Xu and Wang 2009). A further question about boundary marking is whether it also involves other cues. There is some evidence that at least at the phrasal level, no direct F_0 marking is involved in Mandarin (Xu and Wang 2009). At the sentential level, however, the issue is still unclear. Sentence-final lengthening is even more extensive than that at the phrasal level, thus using the same cue as sentence-medial boundaries (Tseng et al. 2005; Yuan et al. 2006). However, such lengthening has also been linked to topic shift (Smith 2004). In other words, there is a conceptual overlap between two functions: topic and boundary. Further research on the issue is clearly needed.

AQ: Please provide complete details of "Xu & Wang, 2009" in the reference list.

36.3 From Individual Effects to Global Patterns

While the full picture about the coexistence of tone and intonation in Chinese languages is still far from clear, an increasingly detailed sketch is emerging from continued empirical research. As the discussion so far has shown, intonational information is encoded by

modifying various aspects of the target approximation process for lexical tones, including, in particular, pitch range and duration. These modifications are associated with various specifically defined communicative functions. Interestingly, once the impacts of these modifications are added together, some of the widely recognized global patterns also start to emerge, including, in particular, declination and rhythm.

36.3.1 Declination

With the exception of final F_0 raising by question modality, most of the pitch range modifications by various functions are in the direction of generating an F_0 down trend over the course of an utterance. First, from the left edge, a *new topic* raises sentence-initial F_0 (by about at least 1 st across different topic levels; Wang and Xu 2011) and lets the subsequent F_0 gradually drop. Second, if the sentence happens to have a nonfinal *focus*, the global F_0 pattern is further tilted in two ways: on-focus raising (unless the tone is Low) and postfocus lowering. Third, the final word, and especially the final syllable of a statement, is lowered in F_0 to contrast with the rising question intonation. Fourth, whenever a sentence consists of more than just the High-level tone (Tone 1 in the case of Mandarin), the non-High tone lowers the F_0 of the following tone relative to the preceding tone. This effect is known as downstep (Stewart 1965). Being directly related to lexical tones, it could be fully attributed to an articulatory effect of carryover assimilation. However, equations [1] and [2] given earlier would predict more than sufficient time for F_0 to fully recover after a non-High tone (because a syllable of 180 ms allows F_0 to shift either up or down by more than 12 st, well over the amount of F_0 shift between any two adjacent tones). The fact that full recovery does not happen suggests that speakers apply insufficient amount of effort to return to the original pitch height. Thus it is possible that downstep is also partially attributable to topic and/or modality.

Putting these pitch range modifications together thus leads to a global F_0 downtrend similar to the intonational phenomenon known as *declination*, which refers to an observed gradual decline in F_0 over the course of an utterance (Cohen and 't Hart 1967). But the knowledge about the different sources of this downtrend allows us to see that declination is likely a byproduct of a joint effect. In Figure 36.3a, for example, we can see only a slight downtrend over the course of a five-syllable all-High-tone sentence. When the sentence consists of alternating High and Low tones, a much larger overall downtrend can be seen. But with the all-High-tone sentence as the reference, we can see clearly that this is due to two downsteps that each renders a post-Low High tone lower than the pre-Low High tone. Taking a closer look at the first downstep, we can also see that it consists of two components: pre-Low F_0 raising and post-low F_0 lowering. While the latter is the familiar carryover effect, the former is known as *anticipatory raising* or *anticipatory dissimilation*, which has been found not only in Mandarin (Xu 1997) but also in Yoruba (Laniran and Clements 2003), Thai (Gandour et al. 1994), and Cantonese (Gu and Lee 2007). In Figure 36.3b, an even greater downtrend can be seen thanks to the on-focus raising of the High tone and postfocus lowering of all subsequent tones.

Together with the topic raising and sentence final-lowering, which are more readily observed in longer sentences (Liu and Xu 2005; Wang and Xu 2011), the global declination is therefore attributable to a combination of individual mechanisms.

36.3.2 Rhythm

A widely known hypothesis is that languages of the world can be divided into three rhythm classes depending on the type of unit involved in manifesting an isochrony tendency: stress-timed, syllable-timed, or mora-timed (Abercrombie 1967). Although empirical research has repeatedly shown that no true isochrony exists (Ramus et al. 1999), more recent efforts have been devoted to developing measurements that can still separate languages along the rhythm-class dimension (Ramus et al. 1999). Like declination, however, rhythm is a global pattern summarized across all the segments in many sentences in a speech corpus. So an obvious question is whether, also like declination, the grossly measured rhythm pattern as a gestalt is decomposable into individual temporal properties, each with a specific articulatory or communicative source. Indeed, some of the sources are plainly obvious. A stress language like English, in which syllable duration is a major acoustic cue of word stress (Fry 1958), undoubtedly would exhibit a gross temporal pattern different from that of a nonstress language like French. A lesser-known source that clearly contributes to an isochrony tendency is known as polysyllabic shortening (Lehiste 1972), which generates a trend toward equal duration for words of different lengths. Interestingly, however, there is evidence that polysyllabic shortening is actually stronger in Mandarin, arguably a syllable-timed language (Lin and Wang 2005) than in English (Nakatani et al., 1981; Xu and Wang 2009). More important, because polysyllabic shortening is directly related to boundary marking (Xu and Wang 2009), it is a byproduct of a clearly defined communicative function (Tseng et al. 2005; Wagner 2005; Yuan et al. 2006). Furthermore, duration patterns are affected by additional factors whose sources are relatively clear: intrinsic segment duration (articulatory; Klatt 1973), focus (functional), and contrastive vowel length (functional). Overall, the evidence suggests that speech rhythm is likely an epiphenomenon derived from a number of independent articulatory and functional mechanisms, and as such it has little to do with either isochrony or language-specific holistic temporal control.

36.4 Summary

To summarize, the concurrent realization of tone and intonation in Chinese languages is likely achieved by controlling different aspects of the target approximation process. While lexical tones are mainly encoded by local pitch targets in terms of height and slope, various intonational functions are likely encoded mainly via modifications of pitch range and syllable duration. Once the contributions of each of the articulatory

mechanisms and functional codes are recognized, it is possible to also better understand global intonational patterns such as declination and rhythm.

References

Abercrombie, D. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Büring, D. 2003. On D-trees, beans, and B-accents. *Linguistics and Philosophy* 26.5:511–545.

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

Cohen, A., and J. 'T Hart. 1967. On the anatomy of intonation. *Lingua* 19:177–192.

Cooper, W.E., S.J. Eady, and P.R. Mueller. 1985. Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America* 77: 2142–2156.

De Jong, K. 2004. Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *Journal of Phonetics* 32:493–516.

Duanmu, S. 1994. Syllabic weight and syllable durations: a correlation between phonology and phonetics. *Phonology* 11:1–24.

Fairbanks, G. 1959. *Voice and Articulation Drillbook*. New York: Harper & Row.

Fry, D B. 1958. Experiments in the perception of stress. *Language and Speech* 1:126–152.

Gandour, J., S. Potisuk, and S. Dechongkit. 1994. Tonal coarticulation in Thai. *Journal of Phonetics* 22:477–492.

Gu, W., and T. Lee. 2007. Effects of tonal context and focus on Cantonese Fo. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 1033–1036. Saarbrücken: Univ. des Saarlandes.

Hirschberg, J. 2000. A corpus-based approach to the study of speaking style. In *Prosody: Theory and Experiment. Studies Presented to Gösta Bruce*, ed. by M. Horne, 335–350. Boston: Kluwer Academic.

Honorof, D.N., and D.H. Whalen. 2005. Perception of pitch location within a speaker's Fo range. *Journal of the Acoustical Society of America* 117:2193–2200.

Klatt, D. H. 1973. Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America* 54:1102–1104.

Laniran, Y.O., and G.N. Clements. 2003. Downstep and high raising: interacting factors in Yoruba tone production. *Journal of Phonetics* 31:203–250.

Ladd, D.R. 2008. *Intonational Phonology*. Cambridge: Cambridge University Press.

Lehiste, I. 1972. The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America* 51:2018–2024.

_____. 1975. The phonetic structure of paragraphs. In *Structure and Process in Speech Perception*, ed. by A. Cohen and S.E.G. Nooteboom, 195–206. New York: Springer-Verlag.

Li, A., Q. Fang, and J. Dang. 2011. Emotional intonation in a tone language: experimental evidence from Chinese. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 1198–1201. Hong Kong: City University.

Liberman, M., and J. Pierrehumbert. 1984. Intonational invariance under changes in pitch range and length. In *Language Sound Structure*, ed. by M. Aronoff and R. Oehrle, 157–233. Cambridge, MA: *MIT Press*.

Lin, H., and Q. Wang. 2005. Mandarin rhythm: an acoustic study. *Journal of Chinese Language and Computing* 17:127–140.

Liu, F., and Y. Xu. 2005. Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica* 62:70–87.

Mueller-Liu, P. 2006. Signalling affect in Mandarin Chinese—the role of utterance-final non-lexical edge tones. In *Proceedings of Speech Prosody 2006*, PS6-3-0048. Dresden: TUD Press.

Nakajima, S., and J.F. Allen. 1993. A study on prosody and discourse structure in cooperative dialogues. *Phonetica* 50:197–210.

Nakatani, L.H., K.D. O'Connor, and C.H. Aston. 1981. Prosodic aspects of American English speech rhythm. *Phonetica* 38:84–106.

Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation*. PhD diss., Massachusetts Institute of Technology, Cambridge.

Prom-On, S., Y. Xu, and B. Thipakorn. 2009. Modeling tone and intonation in Mandarin and English as a process of target approximation. *Journal of the Acoustical Society of America* 125:405–424.

Ramus, F., M. Nespor, and J. Mehlera 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73:265–292.

Sluijter, A.M.C., and V.J. Van Heuven. 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100:2471–2485.

Smith, C.L. 2004. Topic transitions and durational prosody in reading aloud: production and modeling. *Speech Communication* 42:247–270.

Stewart, J. M. 1965. *The Typology of the Twi Tone System*. Legon, Ghana: Institute of African Studies, University of Ghana.

Swerts, M. 1997. Prosodic features at discourse boundaries of different length. *Journal of the Acoustical Society of America* 101:514–521.

Titze, I. R. 1989. On the relation between subglottal pressure and fundamental frequency in phonation. *Journal of the Acoustical Society of America* 85:901–906.

Tseng, C.Y., S.H. Pin, Y.H. Lee, H.-M. Wang, and Y.-C. Chen. 2005. Fluent speech prosody: framework and modeling. *Speech Communication* 46:284–309.

Umeda, N. 1982. "Fo declination" is situation dependent. *Journal of Phonetics* 10:279–290.

Wagner, M. 2005. Prosody and recursion. PhD diss., Massachusetts Institute of Technology, Cambridge.

Wang, B., L. Wang, and T. Kadir. 2011. Prosodic encoding of focus in six languages in China. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 144–147. Hong Kong: City University.

Wang, B., and Y. Xu. 2011. Differential prosodic encoding of topic and focus at sentence initial position in Mandarin Chinese. *Journal of Phonetics* 39:595–611.

Wu, W.L. 2009. Sentence-final particles in Hong Kong Cantonese: are they tonal or intonational? In *Proceedings of Interspeech 2009*, 2291–2294. Grenoble: International Speech Communication Association.

Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25:61–83.

_____. 1999. Effects of tone and focus on the formation and alignment of Fo contours. *Journal of Phonetics* 27:55–105.

_____. 2011. Post-focus compression: cross-linguistic distribution and historical origin. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 152–155. Hong Kong: City University.

Xu, Y., S.-W. Chen, and B. Wang. 2012. Prosodic focus with and without post-focus compression (PFC): a typological divide within the same language family? *The Linguistic Review* 29:131–147.

Xu, Y., and X. Sun. 2002. Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America* 111:1399–1413.

Xu, Y., and Q.E. Wang. 2001. Pitch targets and their realization: evidence from Mandarin Chinese. *Speech Communication* 33:319–337.

Yuan, J. 2004. Intonation in Mandarin Chinese: Acoustics, perception, and computational modeling. Unpublished dissertation, Cornell University, Ithaca, NY.

Yuan, J., M. Liberman, and C. Cieri. 2006. Towards an integrated understanding of speaking rate in conversation. In *Proceedings of Interspeech 2006*, 541–544. Grenoble: International Speech Communication Association.

Zemlin, W.R. 1988. *Speech and Hearing Science—Anatomy and Physiology*. Englewood Cliffs, NJ: Prentice Hall.