Xu, Y. (2017). Intonation. In Encyclopedia of Chinese Language and Linguistics. W. Behr, g. G. Yue, ZevHandel, C.-T. J. Huang and j. Myers. Boston: Brill pp. 458-466.

Intonation in Chinese

Yi Xu

Intonation, when used in a narrow sense, refers to fundamental frequency (henceforth F_0) patterns in speech that convey information beyond lexical meanings. In a broader sense, intonation also includes other suprasegmental patterns such as duration, intensity and voice quality, etc., that also convey non-lexical meanings. This article will focus mainly on the former, because, as will be seen, F_0 is the main carrier of non-lexical meaning, although for boundary marking, duration likely plays a major role.

Function versus form

As the definitions above indicate, there are two sides to intonation, namely, its form and function. The form refers to various observable patterns, while the function refers to the meanings they carry. Like most other aspects of speech, intonation has been studied primarily for its form, i.e., its shape and structure. This is the case with major theories of intonation for European languages, including nuclear tone analysis (O'Connor & Arnold 1961; Wells 2006), the Bolinger (1986, 1989) theory, and the Autosegmental-Metrical (AM) theory (Pierrehumbert 1980; Ladd 2008). For Chinese, the classical work of Chao (1968) lists 13 intonation patterns, all of them primarily defined in terms of form. The recent trend in research, however, is putting increasing emphasis on the functional aspect of intonation (Hirst 2005; Kohler 2005; Xu 2005). This article will reflect this trend, and describe Chinese intonation in terms of how specific communicative functions are encoded with unique intonational properties.

An advantage of studying intonation of Chinese over many other languages is that its tonal nature forces us to discover how it is possible for tone and intonation, which are functionally distinct from each other, to co-exist even though both use F_0 as their primary carrier. An idea made popular by Chao (1968) is that tones are like small ripples riding on large waves of intonation. More recent scientific findings support this idea in general, but also point to specific articulatory and encoding mechanisms that enable the co-production of tone and multiple intonational functions (Xu 2005). This knowledge has increasingly been shown to be applicable not only to Chinese, but also to languages in general (Liu et al. 2013; Prom-on et al. 2009; Xu & Prom-on (submitted)).

Intonational functions

Just as lexical tones serve to distinguish words and morphemes, intonational functions serve to mark contrasts that are meaningful though non-lexical. It is therefore critical to understand intonation by identifying individual functions, their specific encoding schemes, and how they interact with each other. This can be achieved by empirical methods that allow strict control of all relevant factors. The intonational functions that have been recognized this way for Chinese languages include focus, modality (statement/question), topic and boundary marking.

Focus

(1) 他今天上學了 [He went to school today]
(2) 誰今天上學了? [Who went to school?]

(3) 他甚麼時候上學了? [When did he go to school?]
(4) 他今天干甚麼了? [What did he do today]?
(5) 怎麼回事? [What happened?]

When the Mandarin Chinese sentence in (1) is said as a response to different questions as in (2-5), there is an natural emphasis on *tā* 'he', *jīntiān* 'today', *shàngxué le* 'went to school', or none of the words, respectively. Such an emphasis on a particular component of an utterance is known as *prosodic focus*, *narrow focus*, or simply *focus*. Focus in Mandarin has been extensively studied, and the emerging picture is that it is marked by multiple cues. In general, a focused component exhibits expanded pitch range, increased duration and magnified intensity (Chen & Gussenhoven 2008; Chen & Xu 2006; Liu & Xu 2005; Jin 1996; Xu 1999). An example is shown in Figure 1b, where focus is on the first disyllabic word, which clearly contrasts with the neutral focus contours in Figure 1a. Furthermore, the pitch range and intensity of the post-focus region are compressed (narrowed and lowered), a pattern known as *post-focus compression* (PFC), as can be also seen in Figure 1b as compared to Figure 1a. Little, however, is systematically changed in pre-focus words.

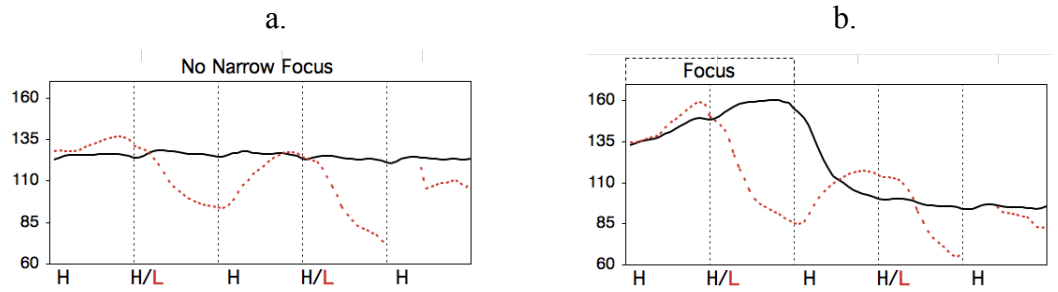

Figure 1. Mean time-normalized F_0 contours of *Māomī mō māomī* 猫咪摸猫咪 [Kitty strokes kitty] (solid line) and *Māomī mō mǎdāo* 猫米摸馬刀 [Kitty strokes saber] (dotted line) with either no narrow focus (a) or focus on the first word (b). Each contour is an average of 40 tokens said by four male speakers of Beijing Mandarin (5 repetitions by each). Adapted from Xu (1999).

The identification of this asymmetrical pitch range, intensity and duration modification by focus in Mandarin (Jin 1996; Xu 1999) as well as in English (Cooper et al. 1985; Xu & Xu 2005) has led to a significant new development in intonation research. That is, PFC, the most stable feature of focus in these languages, turns out to be non-universal, and unevenly distributed among the world's languages, including even among the Sino-Tibetan languages (Xu 2011, Xu et al. 2012). For example, while PFC is present in Běijīng Mandarin, Tibetan, Uygur and Nánchāng 南昌 Gàn 贛, it is absent in Southern Mǐn 閩, Cantonese, Yi, Deang, Wa, and even Táiwān Mandarin (Mandarin spoken in Táiwān) (Wang et al. 2011; Xu et al. 2012). More globally, PFC distribution seems to correlate with language families, e.g., present in Indo-European, Uralic, Altaic and possibly Arabic languages (Hellmuth 2006), but absent in most other language families (Zerbian et al. 2010). This has raised profound questions about the relation between these languages and their evolutionary history. A new area of research is therefore emerging which may lead to a rethinking of language typology, language change and human evolution (Xu 2011).

Modality and utterance-final particles

Modality, also known as *sentence type*, refers to whether an utterance makes a statement or asks a question. This function is often marked morphosyntactically, and in Chinese languages often by sentence-final particles. In addition, however, in Mandarin as in many other languages, a question shows a global rising intonation toward the end of the utterance, whether or not a question particle is present. As found by systematic empirical studies, however, 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 (Liu & Xu 2005; Yuan, 2004). Figure 2 shows mean F_0 contours of Mandarin question versus statement in sentences consisting of only syllables in the first (high level) tone. In Figure 2a focus is either on the sentence-initial word, or there is no narrow focus (hence, neutral focus). In Figure 2b, 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 from the beginning of the sentence or only in the final word of the sentence.

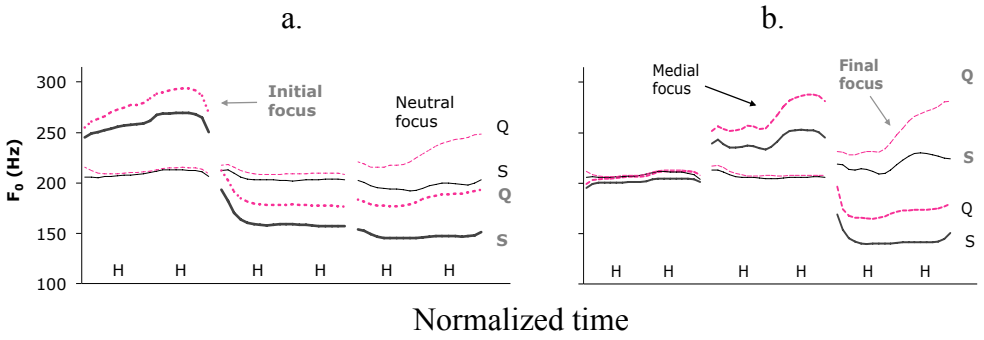

Figure 2. Mean F_0 contours of Mandarin sentence *Zhāng Wēi dānxīn Xiāo Yīng kāichē fāyūn* 張威担心肖英開車發暈 [Zhang Wei is concerned that Xiao Ying may get dizzy when driving] spoken as either a statement or a question. On the left, either focus is on the sentence-initial word (thick lines), or there is no narrow focus (thin lines). On the right, focus is either sentence medial (thick lines) or sentence final (thin lines). The black solid lines represent statements, and the pink dashed lines represent questions. Data from Liu and Xu (2004).

Figure 2 also demonstrates 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 sentence-final rise 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 location of modality coding. This has undermined the notion of a H% boundary tone as a standalone intonational entity (Pierrehumbert 1980), because modality involves adjustments 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, in which the F_0 *lowering* accelerates toward the end of an utterance, resulting in the greatest drop in the final syllable. This final drop is referred to as final lowering or L% boundary tone for English (Liberman & Pierrehumbert 1984). But again, this final drop is likely only part of a continuous F_0 decline that encodes the statement modality.

The accelerating F_0 change found in Mandarin reflects the importance of sentence-final location for marking modality-related information. Beside the question/statement contrast, other information can be also encoded at that location. This is reflected in the wide use of sentence-final particles (SFPs) among the Chinese languages. One of the richest repertoires of SFPs is found in Cantonese, where SFPs make use of a combination of lexical-tone-like F_0 contours and modality-relevant modifications (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 its vocalic component without adding an extra syllable. Similar edge tones have been noted in Mandarin by Chao (1968), who refers to them as particles. Because of the difficulty of eliciting natural edge tones in experimental situations, it is only recently that there has started to be empirical studies of them (Li et al., 2011; Mueller-Liu, 2006). More such research is certainly needed.

Comparison with English

The findings about the encoding of modality in Mandarin and the methodology developed in the research makes it possible to have relatively direct comparisons between Mandarin and other languages. For example, Figure 3 shows that in English F_0 contours of question and statement diverge sharply after focus. In particular, post-focus F_0 in a question rises sharply above the pre-focus level, and then continues to rise gradually all the way to the end of the utterance. This is very different from Mandarin as shown in Figure 2, where post-focus F_0 goes below the pre-focus level even if the utterance is a question.

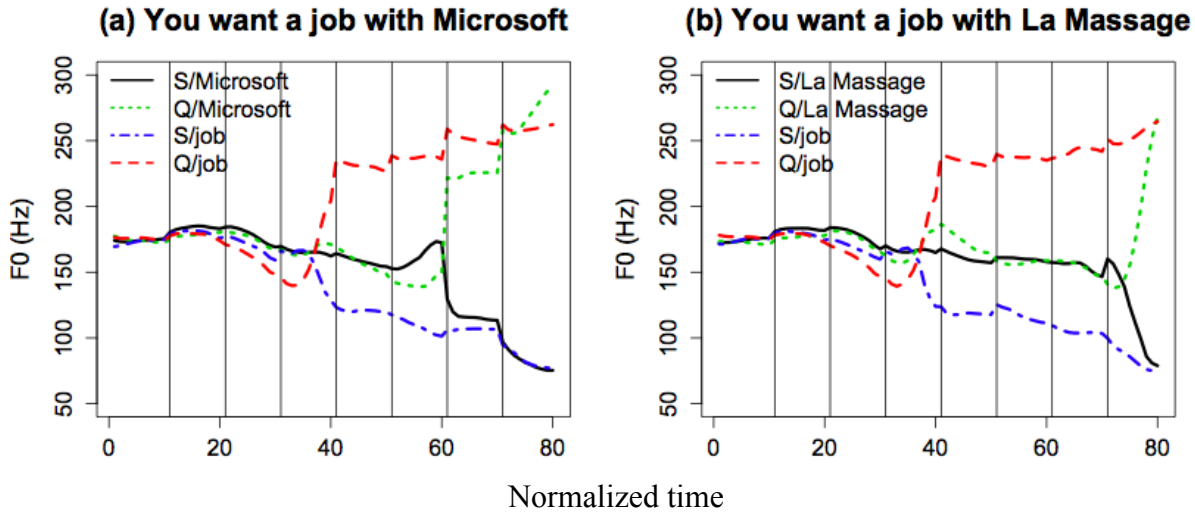

Figure 3. Mean F_0 contours of statements (S) and questions (Q) in American English. The word after "/" is focused. Data from Liu & Xu (2007).

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 higher than in the following sentences (Lehiste 1975; Umeda 1982). This affects the pitch range of the sentence non-linearly, with a large F_0 increase near the beginning and a gradual drop afterwards (Wang & Xu 2011). This gradual drop due to topic is distinct from focus prosody which

immediately lowers post-focus pitch range, as seen in Figure 1b. The communicative function of this topic-related initial F_0 increase 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 & Allen 1993; Wang & Xu 2011), or initiating a new conversational turn (Swerts 1997). Conceptually, topic seems to overlap with some other prosodic notions, in particular 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 is marked by a rising accent in English (Büring 2003). So far, however, no experimental evidence has been reported for such rising accent associated with contrastive topic in English. For Mandarin, a recent study has found no acoustic cue for contrastiveness in a topic (Wang & Xu 2011). The same study also finds that after controlling for focus and topic, newness does not have intonational correlates except for a small duration lengthening.

Boundary marking

The function of boundary marking can be exemplified by an ancient story whose written form is humorously ambiguous in the absence of punctuation:

下	雨	天	留	客	天	天	留	我	不	留
xià	*yǔ*	*tiān*	*liú*	*kè*	*tiān*	*tiān*	*liú*	*wǒ*	*bù*	*liú*
down	rain	sky/day	keep	guest	sky/day	sky/day	keep	1sg	not	keep

Among its possible interpretations are

(a) 下雨天，留客天。天留，我不留。

'A rainy day (is) a day for keeping visitors. The weather keeps the visitor but I don't.'

(b) 下雨天，留客天。天留我不？留！

'A rainy day (is) a day for keeping visitors. Does the weather keep me? Yes!'

In spoken language, expressing meanings equivalent to the differently-punctuated written forms can be achieved by boundary marking, i.e. through pausing and lengthening the domain-final syllable. 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 & Wang 2009). A further question about boundary marking is whether it also involves other cues. There is some evidence that at the phrasal level, no direct F_0 marking is involved in Mandarin (Xu & Wang 2009). At the sentential level, however, the issue is still unclear. Further research on the issue is clearly needed.

From functional intonation patterns to global trends

As the discussion so far has shown, intonational information is encoded by modifying various aspects of the tone production process, 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.

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 downtrend over the course of an utterance. First, from the left edge, a *new topic* raises sentence-initial F_0, and lets the subsequent F_0 gradually drop. Second, whenever a sentence consists of more than just high-level-tone syllables, any non-high-level tone lowers the F_0 of its following tone relative to its preceding tone. In Figure 1a, for example, compared to the slight downtrend over the course of the all-high-tone sentence marked with the solid line (as has also been found by Kochanski & Shih 2003), the sentence consisting of high and low tones marked with the dotted line (tone 1 and tone 3, respectively) shows a much larger overall downtrend. But we can see clearly that this is due to two downsteps which 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 for Mandarin (Xu 1997), but also for Yoruba (Laniran & Clements 2003), Thai (Gandour & Potisok 1994) and Cantonese (Gu & Lee 2007). Second, if the sentence happens to have a non-final *focus*, the global F_0 pattern is further tilted in two ways: on-focus raising (unless the focused tone is Low, e.g., tone 3 in Mandarin) and post-focus 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.

Putting these pitch range modifications together we can see 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 & 't Hart 1967; Shih 1997). But the different sources of this downtrend shown above suggest that declination is likely a byproduct of different meaningful functions. As such it is unlikely to be due to an independent intonation mechanism. A similar find has been made for English by Liberman and Pierrehumbert (1984).

Rhythm

A widely known hypothesis is that languages of the world can be divided into three rhythm classes: stress-timed, syllable-timed and mora-timed (Abercrombie 1967; Ramu et al. 1999; Trubetzkoy 1958). Like declination, the grossly measured rhythm pattern as a gestalt is decomposable into individual temporal properties, each with a specific articulatory or communicative source. An important source of the isochrony tendency is polysyllabic shortening (Lehiste 1972; Nakatani et al. 1981), which generates a trend toward equal duration for words of different lengths. Interestingly, however, there is recent evidence that polysyllabic shortening is stronger in the supposedly syllable-timed Mandarin (Lin & Wang 2005) than in the stress-timed English (Xu & Wang 2009). Because polysyllabic shortening is directly related to boundary marking, it itself is a likely byproduct of a clearly defined communicative function (Tseng et al. 2005; Wagner 2005; Xu & Wang 2009; 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.

Concurrent realization of tonal and intonational functions

All the intonational functions, however, need to be realized concurrently with lexical tones which encode the lexical function (together with consonants and vowels). One way to characterize this process is the parallel encoding and target approximation (PENTA) model (Xu 2005), as shown in Figure 4. The model assumes that communicative functions in speech are conveyed in parallel, as represented by the first block from the left. These functions are parallel to one another, because the meanings they carry are mutually independent rather than governed by a hierarchy. Each function is associated with a unique *encoding scheme*, represented by the second block, which typically consists of multiple properties. These properties control the articulatory parameters represented by the third block. These parameters in turn control an articulatory mechanism of *target approximation* (TA), represented by the fourth block. The TA mechanism, as elaborated in the lower panel of Figure 4, directly generates surface intonation.

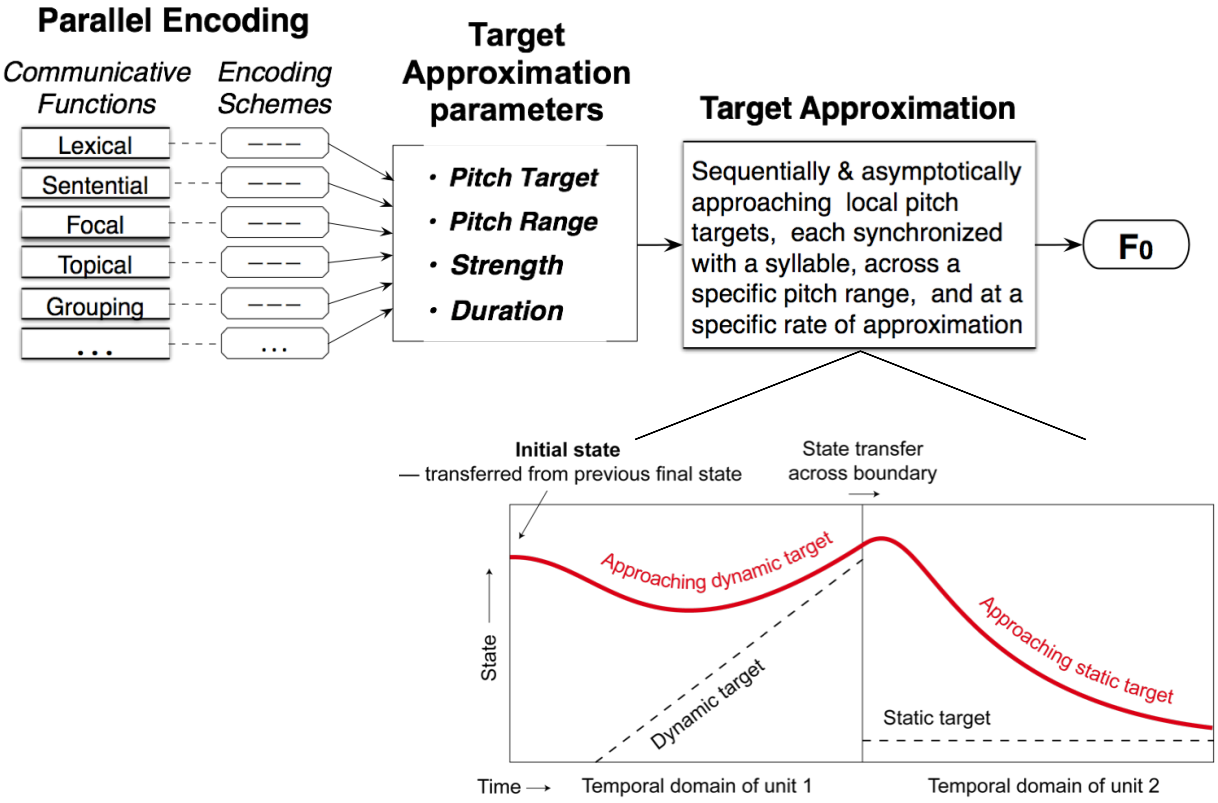

Figure 4. Parallel Encoding and Target Approximation (PENTA) model (Xu, 2005).

The TA model assumes that each syllable is assigned a pitch target that is either static (slope ≈ 0) or dynamic (slope ≠ 0), and surface F_0 results from asymptotic approximation of the underlying target, which are articulated in full synchrony with the syllable. At the boundary between adjacent syllables, the final articulatory state of one syllable is transferred to the next syllable, which often lead to a delay of the apparent alignment of an F_0 turning point, as depicted in the lower panel of Figure 4. Thus all the concurrent tonal and intonational functions jointly determine a linear string of "functionally loaded" local pitch targets, which, when articulated through syllable-synchronized target approximation, generate surface F_0 contours that carry all

the encoded meanings. This conceptual process has been quantified into a computational model that has been tested on Mandarin as well as a number of other languages (Liu et al. 2013; Prom-on et al. 2009; Xu & Prom-on submitted).

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 lexical pitch targets 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. The patterns of intonation described in this article are predominantly based on Mandarin Chinese for which most of the systematic empirical studies have been conducted. But the function of prosodic focus has been investigated in recent years in a number of other Chinese dialects, and a pattern very different from those of Mandarin has been found in some of these dialects. But overall much more research is needed on intonation of Chinese languages and dialects other than Mandarin. The research on Chinese tone and intonation has also led to a computational model that is applicable to both Chinese and non-Chinese languages. Much more research, however, is needed to both collect systematic empirical data from various Chinese languages and perform computational modeling to further enhance our knowledge about Chinese intonation and intonation of languages in general.

References

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

Bolinger, Dwight, *Intonation and its Parts: Melody in Spoken English*, Palo Alto: Stanford University Press, 1986.

Bolinger, Dwight, *Intonation and Its Uses -- Melody in Grammar and Discourse*, Stanford, California: Stanford University Press, 1989.

Büring, Daniel, "On D-Trees, Beans, and B-Accents", *Linguistics and Philosophy* 26, 2003, 511-545.

Chao, Yuen-Ren, *A Grammar of Spoken Chinese*, Berkeley, CA: University of California Press, 1968.

Chen, Yiya. and Carlos Gussenhoven, "Emphasis and tonal implementation in Standard Chinese", *Journal of Phonetics* 36, 2008, 724-746.

Chen, Yiya. and Yi Xu, "Production of weak elements in speech -- Evidence from f0 patterns of neutral tone in standard Chinese", *Phonetica* 63, 2006, 47-75.

Cohen, Antonie and Johan 't Hart, "On the anatomy of intonation", *Lingua* 19, 1967, 177-192.

Cooper, William E., Stephen J. Eady, and Pamela R. Mueller, "Acoustical aspects of contrastive stress in question-answer contexts", *Journal of the Acoustical Society of America* 77, 1985, 2142-2156.

Fry, Dennis B., "Experiments in the perception of stress", *Language and Speech* 1, 1958, 126-152.

Gandour, Jack, Siripong Potisuk and Sumalee Dechongkit, "Tonal coarticulation in Thai", *Journal of Phonetics*, 22, 1994, 477-492.

Gu, Wentao and Lee, Tan, "Effects of tonal context and focus on Cantonese F_0", *Proceedings of The 16th International Congress of Phonetic Sciences*, Saarbrucken, 2007, 1033-1036.

Hellmuth, Sam, "Focus-related pitch range manipulation (and peak alignment effects) in Egyptian Arabic", *Proceedings of Speech Prosody 2006*, Dresden, Germany, 2006, PS4-12-164.

Jin, Shunde, *An Acoustic Study of Sentence Stress in Mandarin Chinese*, Ph.D. dissertation, The Ohio State University, 1996.

Klatt, Dennis H., "Interaction between two factors that influence vowel duration", *Journal of the Acoustical Society of America* 54, 1973, 1102-1104.

Kochanski, Greg, Chilin Shih and Hongyan Jing. Quantitative measurement of prosodic strength in Mandarin. *Speech Communication*, 2003, 41, 625-645.

Laniran, Yetunde O. and G. Nick Clements, "Downstep and high raising: interacting factors in Yoruba tone production", *Journal of Phonetics* 31, 2003, 203-250.

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

Lehiste, Ilse, "The timing of utterances and linguistic boundaries", *Journal of the Acoustical Society of America* 51, 1972, 2018-2024.

Lehiste, Ilse, "The phonetic structure of paragraphs", in: A. Cohen and S. E. G. Nooteboom, eds., *Structure and Process in Speech perception*, New York: Springer-Verlag, 1975, 195-206.

Li, Aijun, Qiang Fang, and Jianwu Dang, "Emotional intonation in a tone language: Experimental evidence from Chinese", *Proceedings of The 17th International Congress of Phonetic Sciences*, Hong Kong**:** 2011, 1198-1201.

Liberman, Mark and Janet Pierrehumbert, "Intonational invariance under changes in pitch range and length", in: M. Aronoff and R. Oehrle, eds., *Language Sound Structure*, Cambridge, Massachusetts: M.I.T. Press, 1984, 157-233.

Lin, Hua and Qian Wang, "Mandarin rhythm: an acoustic study", *Journal of Chinese Language and Computing* 17, 2007, 127-140.

Liu, Fang. "Single vs. double focus in English statements and yes/no questions" *Proceedings of Speech Prosody 2010*, Chicago, 2010.

Liu, Fang and Yi Xu, "Parallel encoding of focus and interrogative meaning in Mandarin intonation", *Phonetica* 62, 2005, 70-87.

Liu, Fang, Yi Xu, Santitham Prom-on and Alan C. L. Yu. "Morpheme-like prosodic functions: Evidence from acoustic analysis and computational modeling". *Journal of Speech Sciences*, 2013, 3, 85-140.

Mueller-Liu, Patricia, “Signalling affect in Mandarin Chinese - the role of utterance-final non-lexical edge tones”, *Proceedings of Speech Prosody* 2006, Dresden, Germany**,** 2006, PS6-3-0048.

Nakajima, Shin'ya and James F. Allen, “A study on prosody and discourse structure in cooperative dialogues”, *Phonetica* 50, 1993, 197-210.

Nakatani, Lloyd H., Kathleen D. O'Connor and Carletta H. Aston, “Prosodic aspects of American English speech rhythm”, *Phonetica* 38, 1981, 84-106.

O'Connor, Joseph D. and Gordon F. Arnold, *Intonation of Colloquial English*, London: Longmans, 1961.

Pierrehumbert, Janet, *The Phonology and Phonetics of English Intonation*, Ph.D. dissertation, MIT, Cambridge, MA, 1980.

Prom-on, Santitham and Yi Xu. “Modeling speech melody as communicative functions with PENTAtrainer2” *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France, 2013. 82-85.

Prom-on, Santitham, Yi Xu and Bundit Thipakorn. “Modeling tone and intonation in Mandarin and English as a process of target approximation”. *Journal of the Acoustical Society of America*, 2009, 125, 405-424.

Ramus, F., Marina Nesporb and Jacques Mehlera, “Correlates of linguistic rhythm in the speech signal”, *Cognition* 73, 1999, 265-292.

Swerts, Marc, “Prosodic features at discourse boundaries of different length”, *Journal of the Acoustical Society of America* 101, 1997, 514-521.

Tseng, Chiu-yu, Shao-huang Pin, Yehlin Lee, Hsin-min Wang and Yong-cheng Chen, “Fluent speech prosody: Framework and modeling”, *Speech Communication* 46, 2005, 284-309.

Trubetzkoy, Nikolai S., *Grundzuüge der Phonologie*, Göttingen: Vandenhoeck & Ruprecht, 1958.

Umeda, Noriko, “‘F0 declination’ is situation dependent”, *Journal of Phonetics* 10, 1982, 279-290.

Wagner, Michael, *Prosody and Recursion*, Ph.D. Dissertation, Massachusetts Institute of Technology, 2005.

Wang, Bei, Ling Wang, and Tursun Kadir, “Prosodic encoding of focus in six languages in China”, *Proceedings of The 17th International Congress of Phonetic Sciences*, Hong Kong, 2011, 144-147.

Wang, Bei and Yi Xu, “Differential prosodic encoding of topic and focus at sentence initial position in Mandarin Chinese”, *Journal of Phonetics* 39, 2011, 595-611.

Wells, John C., *English Intonation: An Introduction*, Cambridge: Cambridge University Press, 2006.

Wu, Wing Li, “Sentence-final particles in Hong Kong Cantonese: Are they tonal or intonational?”, *Proceedings of Interspeech 2009*, 2291-2294.

Xu, Yi, “Effects of tone and focus on the formation and alignment of F0 contours”, *Journal of Phonetics* 27, 1999, 55-105.

Xu, Yi “Speech melody as articulatorily implemented communicative functions”. *Speech Communication*, 2005, 46, 220-251.

Xu, Yi, “Post-focus compression: Cross-linguistic distribution and historical origin”, *Proceedings of The 17th International Congress of Phonetic Sciences*, Hong Kong, 2011, 152-155.

Xu, Y. and C. Xueqing Xu, “Phonetic realization of focus in English declarative intonation”, *Journal of Phonetics* 33, 2005, 159-197.

Xu, Yi and Maolin Wang, “Organizing syllables into groups—Evidence from F0 and duration patterns in Mandarin”, *Journal of Phonetics* 37, 2009, 502-520.

Xu, Yi and Santitham Prom-on. “Toward invariant functional representations of variable surface F0 contours: Synthesizing speech melody via model-based stochastic learning”, *Submitted*.

Xu, Yi, Szu-wei Chen and Bei Wang, “Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family?”, *The Linguistic Review* 29, 2012, 131-147.

Yuan, Jiahong, *Intonation in Mandarin Chinese: Acoustics, Perception, and Computational Modeling*, PhD dissertation, Cornell University, Ithaca, NY, 2004.

Yuan, Jiahong, Mark Liberman and Christopher Cieri, “Towards an Integrated Understanding of Speaking Rate in Conversation”, *Proceedings of Interspeech 2006*, 2006, 541-544.

Zerbian, Sabine, Susanne Genzel and Frank Kügler, “Experimental work on prosodically-marked information structure in selected African languages (Afroasiatic and Niger-Congo)”, *Proceedings of Speech Prosody 2010*, Chicago, 2010.