In Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing. pp. 215-220.

Transmitting Tone and Intonation Simultaneously — The Parallel Encoding and Target Approximation (PENTA) Model

Yi Xu

Haskins Laboratories
New Haven, CT, USA
xu@haskins.yale.edu

Abstract

Lexical tones use F_0 to distinguish between words that are otherwise phonemically identical. Intonation uses F_0 to convey discourse, attitudinal and affective information that is often not directly encoded in the words or syntax of the spoken utterances. Because the same acoustic parameter is being used, it is a question how well lexical tones and intonation can coexist in a language. The Parallel Encoding and Target Approximation (PENTA) model proposed here addresses the concurrent transmission of these two aspects of speech in a principled way. As will be demonstrated, the PENTA model can help us better understand both the capacity and limitation of F_0 contours in simultaneously conveying meanings carried by tone and intonation.

1. Introduction

It has been long debated whether intonation components are linearly sequenced or superposed [15, 18, 23, 32-34]. For a tone language one may argue that melodic components of speech have to be superpotional. In a Mandarin utterance, for example, all syllables are specified with lexical tones (including the neutral tone). Meanwhile, the utterance can be either a question or a statement; one part of it may be focused; the utterance could also carry a new topic or initiate a conversational turn. Finally, there could be demarcation information also manifested partly through F_0. All these communicative functions are transmitted virtually simultaneously. Thus to a first approximation, F_0 formation has to be superpositional. On the other hand, as has been forcefully argued, straight superposition models are not flexible enough to account for all the observed F_0 variations, because various components of intonation are not simply added together on a linear or nonlinear scale [23, 33]. In fact, recent advances in research on tone and intonation suggest that a comprehensive model of speech melody has to satisfy at least three critical requirements. First, it has to make a clear separation between the meaning-bearing components of intonation, which are *functionally defined*, and the primitives of speech melody, which are *defined purely in forms* (i.e., detached from meanings) and are readily implementable in articulation. Second, it has to specify a continuous link between articulatory mechanisms of F_0 contour generation and the functional components of speech melody. Third, it has to specify mechanisms for concurrent transmission of tonal and *multiple* intonational functions. In this paper, I propose the Parallel Encoding and Target Approximation (PENTA) model that attends to all of these requirements. As I will show, the PENTA model provides a framework in which a rich repertoire of communicative functions can be realized concurrently through F_0, with all the details of the F_0 contours still linked to their proper sources.

2. The PENTA Model

The PENTA model is based on two basic assumptions. First, speech melody is produced by an articulatory system whose physical and neurological properties impose various constraints on the way speech melody is generated. Second, a multitude of communicative functions are concurrently conveyed through speech melody and perceptual parsing of these functions requires that each function be *uniquely* encoded. Following the two basic assumptions and drawing upon findings from recent research, three primary hypotheses are incorporated into the PENTA model. (1) The communicative functions are encoded in parallel by specifying the values of the melodic primitives using distinctive encoding

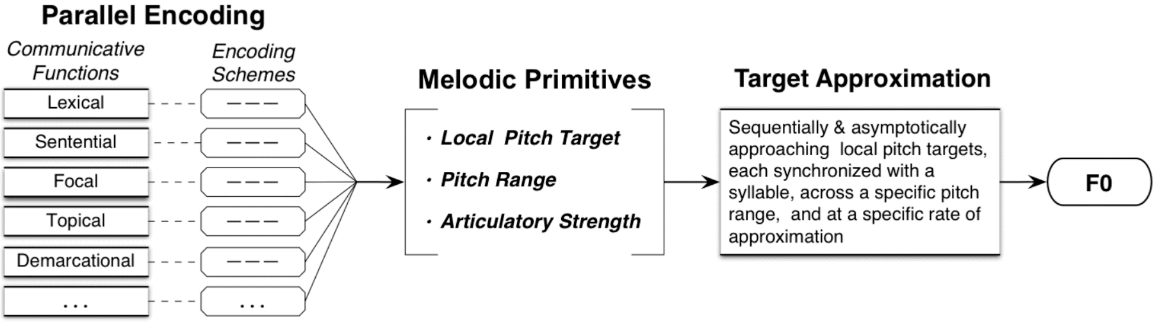

Figure 1. *A sketch of the PENTA model. See text for explanations. The unnamed block at the bottom left indicates communicative functions yet to be explored.*

schemes. (2) The melodic primitives are in the form of *local pitch target*, *pitch range* and *articulatory strength*, which are at once basic encoding elements for the communicative functions and control parameters for the articulatory system that generates F_0 contours. (3) Taking the primitives as input, the articulatory system generates F_0 by successively approaching local pitch targets, each synchronized with a syllable, across specific pitch ranges, and with specific articulatory strengths.

A diagram of the PENTA model is shown in Fig. 1. The stacked boxes on the far left represent individual communicative functions to be conveyed through melody. As can be seen, lexical tone is only one of them. All these functions control F_0 through encoding schemes (shown to their right) that specify the values of the melodic primitives, including local pitch target, pitch range and articulatory strength. These encoding schemes thus form an arbitrary (i.e., conventional or stipulative) link between meaningful communicative functions and meaningless melodic primitives. Table□ shows possible symbolic values of the parameters and their notational representations. Local pitch targets are defined by two parameters: height and slope. Height specifies the relative frequency level of the target within the scope defined by pitch range, and slope specifies whether the target is static or dynamic. When slope is zero, as in the case of [high], [low] or [mid], the target is static. When the slope is positive or negative, as in the case of [rise] or [fall], the target is dynamic. Pitch range specifies the frequency range across which local pitch targets are implemented. It is defined by two parameters: height and span. Height specifies the relative height of the pitch range, e.g., |high|. |low| or |mid|. Span specifies the width of the pitch range, e.g., |wide| or |narrow|. Articulatory strength specifies the speed at which a local pitch target is approached. When articulatory strength is <strong>, the target is approached faster than when it is <weak>.

Table 1. *Possible symbolic values of the melodic primitives: local target, pitch range, and articulatory strength, which may be notationally distinguished from one another by using [ ], %, | | and < >, respectively.*

Local Target:		
	Regular target:	[high], [low], [rise], [fall], [mid]
	Boundary tone:	high%, low%, mid%
Pitch Range:		
	Height:	|high|, |low|, |mid|
	Span:	|wide|, |narrow|, |normal|
Articulatory Strength:		<strong>, <weak>, <normal>

The melodic primitives are, *at the same time*, control parameters for the Target Approximation model [53]. Specifically, the symbolic values shown in Table 1 correspond *directly* to quantitative parameters of the Target Approximation model that simulates articulatory implementation of the local targets across the specified pitch ranges and with specified articulatory strengths. A sketch of the Target Approximation model is shown in Fig. 2.[1]

Fig. 3 illustrates the operation of Parallel Encoding of tone and focus in Mandarin and how they are realized as surface F_0 contours through Target Approximation. Displayed in the graphic part of the figure are the mean F_0 curves of the Mandarin sentence “Māomǐ mō māomī” (tone: HLHHH) said with and without initial focus (thick-solid/-thin curves), together with an all-H sentence as reference (dotted curve). The syllable boundaries are indicated by the vertical lines. The symbolic values of the local pitch targets associated with the lexical tones (block letters in the graph) and pitch ranges associated with the focus are shown below the F_0 plots in two separate tiers. The height and shape of the local pitch targets corresponding to the lexical tones are depicted by the short horizontal lines. The pitch range adjustments by focus corresponding to the pitch range categories are indicated by the block arrows. The two unfilled block arrows on the left indicate pitch range widening directly under focus. The filled block arrow on the right indicates pitch range narrowing and lowering after focus (though the narrowing is not obvious because the tonal targets are static).

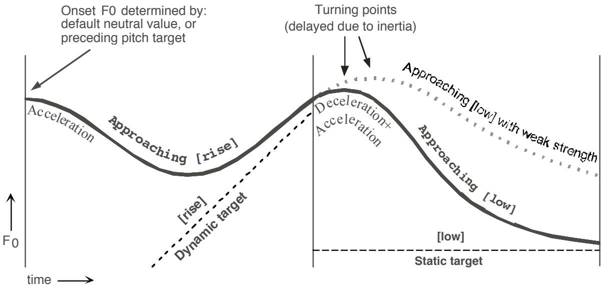

Figure 2. *Illustration of the Target Approximation model. The vertical lines represent syllable boundaries. The straight dashed lines represent underlying* pitch targets. *The solid curve represents the F_0 contour that results from* asymptotic approximation *of the pitch targets. (Adapted from [53].) The dotted curve in syllable 2 simulates the effect of <weak> articulatory strength.*

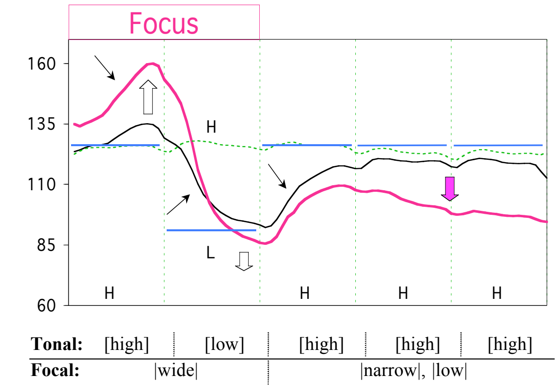

Figure 3. *Averages F_0 of 20 repetitions of the* Mandarin *sentence “Māomǐ mō māomī” [Cat-rice strokes Kitty], and the Mandarin sentence “Māomī mō māomī” [Kitty strokes Kitty], spoken by 4 male speakers. Thick solid curve: focus on “Māomǐ”; thin solid curve: no focus; dashed curve: HHHHH with no focus. Vertical grids indicate locations of nasal murmur onset. (Data from [50].) Short horizontal lines indicate hypothetical pitch targets [high] and [low] [53]. Thin arrows point to F_0 variations due to inertia. Unfilled block arrows indicate on-focus pitch range expansion. Filled block arrow indicates post-focus pitch range narrowing and lowering.*

The quantitative parameters corresponding to the symbolic values of the melodic primitives are turned into continuous F_0 contours through Target Approximation (cf. Fig. 2). Under the Target Approximation model F_0 in each syllable approaches the assigned target asymptotically, giving rise to the extensive transitions during syllables 1-3 in Fig. 3, as indicated by the thin arrows. It also produces the peaks in syllables 1 and 3 (the latter only when with initial focus), and the valley in syllable 3. Also seen in Fig. 3 are the mechanical effects of downstep brought about by L, which raises F_0 of the preceding H and lowers the F_0 of the following H (most obvious in the thin solid curve where the F_0 lowering after L cannot be attributed to an early focus).

Note that the symbolic representations do not necessarily imply categorical degrees of adjustment. Adjustment can be gradient. For example, to initiate a new topic or a conversational turn, a gradient amount of pitch range raising may be used, depending on the newness of the topic or eagerness of the turn-taking, and on the value of the preceding pitch range [41]. Conceivably, the categoricalness of an encoding scheme is dependent on the nature of the specific communicative function.

The PENTA model thus satisfies the three critical requirements laid out in the Introduction: a) separation of meaning and melodic primitives, b) continuous link between functional components and articulation, and c) capacity for multiple communicative functions to be concurrently transmitted. By complying with these requirements, the PENTA model differs from existing models in which the basic components are defined in forms that are immediately meaningful: nucleus, head and tail [6-7, 29, 31], H and L tones manifesting as F_0 peaks and valleys [23, 32-34], accent and phase commands [15], or complex F_0 shapes that are either fully overt or stylized [2, 42-43]. Furthermore, linearity and superposition (in the broad sense) coexist seamlessly in the PENTA model: The implementation of local pitch targets is strictly linear, while parallel encoding allows *multiple* layers of functional meanings to be transmitted concurrently.

The PENTA model is based on extensive experimental data. Evidence for the existence of both dynamic and static local pitch targets has been shown for Mandarin [20, 48-51]. But there has also been both direct and indirect evidence for local targets in other languages such as English [24, 39, 54]. There has been considerable evidence in several languages for multi-region pitch range manipulation by focus [5, 16, 21, 35-36, 50, 54-55] and for pitch range raising by new topic [25, 28, 41, 44]. Articulatory strength as a melodic primitive is evidenced by findings about the neutral tone in Mandarin [4] and unstressed syllable in English [54]. More detailed discussion of the theoretical considerations leading to the PENTA model can be found in [52].

3. Complexity in encoding schemes

Probably because there are so many different communicative functions to be concurrently transmitted through F_0, the uniqueness of an encoding scheme is not always achieved by a single cue. Rather, it often takes a collection of cues to make one function distinct from others. For example, in several languages, at least, focus has been found to be manifested not only by expanding the pitch range of the focused item, but also by compressing the pitch range of post-focus items [19, 21, 35-36, 50], as can be clearly seen in Fig. 3 for Mandarin. New topic, in contrast, probably only raises (i.e., without downwardly expanding) the pitch range of the first non-stressed word in an utterance, while leaving the pitch range of the later words intact [25, 28, 41, 44]. At the same time, lexical stress in languages like English, which has been long known to be partially manifested through F_0 [13], may also be encoded in the melodic specifications of not only the stressed syllables, but also unstressed syllables. This general idea is captured by metrical phonology [17, 26]. But phonetic research on stress typically focuses only on the characteristics of the stressed syllables (e.g., [8, 10, 12-13, 30]). Even lexical tones are not always encoded by local pitch targets alone. The neutral tone in Mandarin, for example, is likely associated with not only a static [mid] or [lower-mid] target, but also a <weak> articulatory strength which makes the approximation of the target much slower than that of the full tone targets [4]. From the perspective of encoding, the weak strength actually helps to make the neutral tone stand out from the rest of the tones.

Note that complexity in encoding schemes does not necessarily mean that all perceptual cues are part of the encoding schemes. For example, although it has been shown that amplitude profile of isolated syllable can help listeners identify lexical tones when F_0 information is missing [14, 46], it is unlikely that speakers deliberately control amplitude when producing the tones. It is more likely that the physiology and physics of F_0 production naturally make amplitude co-vary with F_0, and speakers would have to make a special effort not to let that happen. In contrast, post-focus pitch range compression does not seem to be physically or physiologically obligatory, and hence it is more likely to be an intrinsic part of the coding scheme.

4. Competition between functions

Uniqueness of encoding does not mean that conflict of cues can be always avoided. In fact, with so many functions to be conveyed through F_0, the coding space is quite crowded and competitions among the functions are inevitable. While not yet fully established, a number of phenomena could be related to conflicts between certain functions. The first case involves focus and topic. It has been reported that in a long sentence, an initial focus does not raise the initial F_0 because it is already quite high without it [9]. It has also been found that there is an upper limit on the amount of pitch raising by focus [3]. These findings suggest that the amount of F_0 raising by a new topic is positively related to sentence length, and beyond a certain length, the amount of pitch rising by a new topic may exceed that by an initial focus. Despite this conflict, however, the two functions may still remain distinct from each other in terms of the pitch range of the later words in the sentence. As mentioned earlier, while focus compresses (and lowers, in the case of statements) the pitch range of the post-focus words, a new topic leaves the pitch range of the later words intact.

Another case of potential conflict is between focus and sentence type. In fact, two kinds of conflicts are involved. A glimpse of both can be seen in Fig. 4, which displays mean F_0 contours of statements and questions in English (a) and Mandarin (b, c) with medial, final or no narrow focus. The sentence types and their focus conditions are explained in the figure caption. The first kind of conflict has to do with the realization of final focus. It has been shown in both Mandarin and English that a final focus is not manifested as distinctly from neutral focus as an earlier focus [21, 50, 54-55]. A possible cause of such “encoding deficiency” can be seen in Fig. 4. That is, the final position is where an extra F_0 rise is produced in a question, which has been attributed to a boundary tone [32] (also see work by Shih and by Myers, both in this volume). The extra raising occurs in both English and Mandarin, as can be seen in sentence 5 in all three plots in Fig. 4. As a result, there does not seem to be much room left for a final focus (sentence 3) to raise F_0 without causing it to be confusable with a question. Note in particular that for both speakers shown in Fig. 4, the total pitch range covered by the sentences is about two octaves, which has about exhausted the total pitch range reported elsewhere [11].

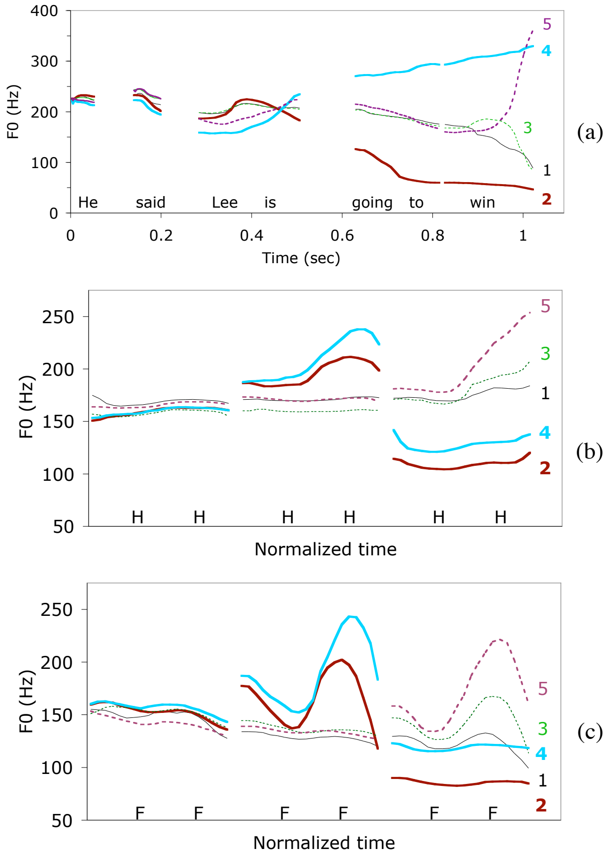

Figure 4. *Mean F_0 contours of statements and questions with middle, final or no focus. (a) “He said Lee is going to win” produced by a female American English speaker (average of 4 repetitions). (b) “Zhāng Wēi (dānxīn) Xiāo Yīng (kāichē) fāyūn” [Zhang We is worried that Xiaoyin will get dizzy while driving] and (c) “Yè Liàng (hàipà) Zhào Lì (shuìjiào) zuòmèng” [Ye Liang is afraid that Zhao Li will dream while sleeping] produced by a male Mandarin speaker (average of 5 repetitions). (Data from [27]) The numbers on the right of the curves indicate sentence type and focus location:*

1: statement with no narrow focus,
2: statement with medial focus,
3: statement with final focus,
4: question with medial focus,
5: question with final focus.

The second kind of conflict is seen in the difference between Mandarin and English regarding post-focus F_0 patterns. As shown in Fig. 4, there are strong interactions between focus and sentence type in both languages, but the interaction seems much smaller in Mandarin than in English. In English, a non-final focus is like a pivot after which F_0 diverges dramatically depending on whether the sentence is a statement or a question, as is evident in sentences 2 and 4. In other words, of the two post-focus F_0 variations typically seen in a statement, i.e., *compression* and *lowering* [50, 54], only the former is uniquely attributable to focus. In sentence 4 in Fig. 4a, while also compressed after focus, the pitch range is raised rather than lowered as in sentence 2. However, the pitch range changes in the Mandarin questions in Fig. 4b and 5c are quite different. Although focus again serves as a pivot point at which statement and question contours start to diverge, post-focus pitch range is lowered rather than raised as in the English questions shown in Fig. 4a. A likely reason for the difference between the two languages is that local pitch targets in Mandarin are lexical, and are not altered by sentential functions (which may not be true with all tone languages), whereas in English, on-focus and post-focus pitch targets seem to be determined jointly by focus and sentence type: [fall] vs. [rise] on-focus, and [low] vs. [high] post-focus in statements vs. questions. It could also be the case that tonal space is more crowded in Mandarin than in English. But there is initial evidence that English also has a rich repertoire of local targets: [high], [low], [mid], [fall] ([54] for General American dialect). It would be interesting to find out whether this difference between the two languages is reflected in the perceptual distinctness of question vs. statement.

Yet another potential conflict is between focus and demarcation. The demarcational function largely corresponds to what is conventionally known as rhythm or prosodic structure, but it is function-based in the sense that it serves to help organize a string of syllables into chunks. The demarcation function is often interfered with by other functions such as focus. As has been argued [1], focus location is largely pragmatically determined and is thus not bound by natural word or phrase boundaries. As also has been demonstrated, focus may be placed on a unit as small as a segment [45]. Thus when the location of focus happens to break a natural unit such as a word or a phrase, the demarcation information may be masked by focus, given that the focus cues are likely to be stronger than the demarcation cues, because it is mainly encoded non-lexically, whereas melodic demarcation information is only auxiliary to many other cues for the same function. A potential consequence of this conflict is that, when it comes to perceptually judging where and how big a break is, as is done in ToBI-style transcriptions [40], the demarcation information may not be clearly separated from the focus information. Thus focus may invoke a clear break percept when the listener’s task during transcription is to judge only the location and size of the breaks.

5. Conclusions

Though still rather rudimentary, the Parallel Encoding and Target Approximation (PENTA) model proposed in this paper allows for both clear separation and smooth integration of the tonal and intonational aspects of speech melody, and specifies a continuous link between articulatory mechanisms of F_0 contour generation and the functional components of speech melody. It also allows for seamless coexistence of linearity and superposition (in the broad sense): The implementation of local pitch targets is strictly linear, while parallel encoding allows *multiple* layers of functional meanings to be transmitted concurrently.

Initial effort to quantify the Target Approximation part of the PENTA model was made in [47]. Effort to quantify the entire model has started, and its progress can be seen in an interactive Java program accessible at

http://www.haskins.yale.edu/yixu/f0_model.html.

As final note, the PENTA model in its current form addresses only F_0 variations in speech. To expand it into a comprehensive model of tone and intonation, other suprasegmentals such as voice quality, duration and amplitude need to be included. Also, in the long run, the conceptual framework represented by the PENTA model is applicable not only to tone and intonation, but also to other aspects of speech, including consonants and vowels and their variations due to prosody.

[1.] Note that the Target Approximation model is in some ways similar to the soft template model (Kochanski & Shih 2003), which describes F_0 contours as resulting from implementing underlying templates with different amounts of muscle forces under the physical constraint of *smoothness*. The two models differs in that, instead of assuming *bidirectional* smoothing as the major constraining mechanism, the Target Approximation model assumes asymmetrical contextual influences: Assimilatory carryover influence, but dissimilatory anticipatory influence.

Acknowledgment

I would like to thank Doug Honorof for his helpful comments. This study was supported in part by NIH grant DC03902.

6. References

[1] Bolinger, D. L., 1972. Accent is predictable (if you're a mind reader). *Language* 48: 633-644.

[2] Bolinger, D., 1989. *Intonation and Its Uses -- Melody in Grammar and Discourse*. Stanford, California: Stanford University Press.

[3] Chen, Y., 2003. *The Phonetics and Phonology of Contrastive Focus in Standard Chinese*. Ph.D. dissertation. State University of New York at Stony Brook.

[4] Chen, Y.; Xu, Y., 2002. Pitch Target of Mandarin Neutral Tone. Presented at LabPhon 8, New Haven, CT. New Haven, CT.

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

[6] Cruttenden, A., 1997. *Intonation*. Cambridge: Cambridge University Press.

[7] Crystal, D., 1969. *Prosodic Systems and Intonation in English*. London: Cambridge University Press.

[8] de Jong, K. J., 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America* 97: 491-504.

[9] Eady, S. J.; Cooper, W. E., 1986. Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America* 80: 402-416.

[10] Engstrand, O., 1988. Articulatory correlates of stress and speaking rate in Swedish. *Journal of the Acoustical Society of America* 83: 1863-1875.

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

[12] Fant, G.; Kruchenberg, A., 1994. Notes on stress and word accent in Swedish. *STL-QPSR* (2-3): 125-144.

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

[14] Fu, Q.-J.; Zeng, F.-G.; Shannon, R. V.; Soli, S. D., 1998. Importance of tonal envelope cues in Chinese speech recognition. *Journal of the Acoustical Society of America* 104: 505-510.

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

[16] Gårding, E., 1987. Speech act and tonal pattern in Standard Chinese. *Phonetica* 44: 13-29.

[17] Goldsmith, J. A., 1990. *Autosegmental and Metrical Phonology*. Oxford: Blackwell Publishers.

[18] Grønnum, N., 1995. Superposition and subordination in intonation — a non-linear approach. In *Proceedings of The 13th International Congress of Phonetic Sciences*, Stockholm. pp. 124-131.

[19] Hasegawa, Y.; Hata, K., 1992. Fundamental frequency as an acoustic cue to accent perception. *Language and Speech* 35: 87-98.

[20] Howie, J. M., 1976. *Acoustical Studies of Mandarin Vowels and Tones*. London: Cambridge University Press.

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

[22] Kochanski, G.; Shih, C., 2003. Prosody modeling with soft templates. *Speech Communication* 39: 311–352.

[23] Ladd, D. R., 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

[24] Ladd, D. R.; Schepman, A., 2003. "Sagging transitions" between high pitch accents in English: experimental evidence. *Journal of Phonetics* 31: 81–112.

[25] Lehiste, I., 1975. The phonetic structure of paragraphs. In *Structure and process in speech perception*. A. Cohen and S. E. G. Nooteboom (eds.). Springer-Verlag: New York: 195-206.

[26] Liberman, M.; Prince, A., 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8: 249-336.

[27] Liu, F. forthcoming. Interaction of sentence type and focus.

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

[29] O'Connor, J. D.; Arnold, G. F., 1961. *Intonation of Colloquial English*. London: Longmans.

[30] Ohala, J. J., 1977. The physiology of stress. In *Studies in Stress and Accent*. L. M. Hyman. (eds.) Los Angeles, CA: Department of Linguistics, University of Southern California: 145-168.

[31] Palmer, H. E., 1922. English Intonation, with systematic exercises. Cambridge: Heffer.

[32] Pierrehumbert, J., 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation. MIT, Cambridge, MA.

[33] Pierrehumbert, J., 2000. Tonal elements and their alignment. In *Prosody: Theory and Experiment — Studies Presented to Gösta Bruce.* M. Horne (eds.). London: Kluwer Academic Publishers: 11-36.

[34] Pierrehumbert, J.; Beckman, M., 1988. *Japanese Tone Structure*. Cambridge, MA: The MIT Press.

[35] Rump, H. H.; Collier, R., 1996. Focus conditions and the prominence of pitch-accented syllables. *Language and Speech* 39: 1-17.

[36] Selkirk, E.; Shen, T., 1990. Prosodic domains in Shanghai Chinese. In *The Phonology-Syntax Connection.* S. Inkelas and D. Zec (eds.). Chicago: University of Chicago Press: 313-37.

[37] Shen, J., 1994. Hanyu yudiao gouzao he yudiao leixing [Intonation structures and patterns in Mandarin]. *Zhongguo Yuwen [Journal of Chinese Linguistics]* 1994-3: 221-228.

[38] Shih, C.-L., 1988. Tone and intonation in Mandarin. *Working Papers, Cornell Phonetics Laboratory* (No. 3): 83-109.

[39] Silverman, K. E. A.; Pierrehumbert, J. B., 1990. The timing of prenuclear high accents in English. In *Papers in Laboratory Phonology 1 — Between the Grammar and Physics of Speech.* J. Kingston and M. E. Beckman (eds.). Cambridge: Cambridge University Press: 72-106.

[40] Silverman, K.; Beckman, M.; Pitrelli, J.; Ostendorf, M.; Wightman, C.; Price, P.; Pierrehumbert, J.; Hirschberg, J., 1992. *ToBI: A standard for labeling English prosody.* In *Proceedings of The 1992 International Conference on Spoken Language Processing*, Banff. pp. 867-870.

[41] Swerts, M.; Ostendorf, M., 1997. Prosodic and lexical indications of discourse structure in human-machine interactions. *Speech Communication* 22: 25-41.

[42] 't Hart, J.; Collier, R.; Cohen, A., 1990. A perceptual Study of Intonation — An experimental-phonetic approach to speech melody. Cambridge: Cambridge University Press.

[43] Taylor, P., 2000. Analysis and synthesis of intonation using the Tilt model. *Journal of the Acoustical Society of America* 107: 1697-1714.

[44] Umeda, N., 1982. "F0 declination" is situation dependent. *Journal of Phonetics* 10: 279-290.

[45] van Heuven, V. J., 1994. What is the smallest prosodic domain? In *Papers in Laboratory Phonology*. P. A. Keating. Cambridge University Press. 3: 76-98.

[46] Whalen, D. H.; Xu, Y., 1992. Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica* 49: 25-47.

[47] Xu, C. X.; Xu, Y.; Luo, L.-S., 1999. A pitch target approximation model for F0 contours in Mandarin. In *Proceedings of The 14th International Congress of Phonetic Sciences*, San Francisco. pp. 2359-2362.

[48] Xu, Y., 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25: 61-83.

[49] Xu, Y., 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55: 179-203.

[50] Xu, Y., 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 27: 55-105.

[51] Xu, Y., 2001. Fundamental frequency peak delay in Mandarin. *Phonetica* 58: 26-52.

[52] Xu, Y., 2004. Separation of functional components of tone and intonation from observed F0 patterns. To appear in *Festschrift in Honor of Zongji Wu.*

[53] Xu, Y.; Wang, Q. E., 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33: 319-337.

[54] Xu, Y.; Xu, C. X., forthcoming. Intonation components in short English statements.

[55] Xu, Y.; Xu, C. X.; Sun, X., 2004. On the Temporal Domain of Focus. To appear in *Proceedings of The 2nd International Conference on Speech Prosody*, Nara, Japan.