Intonation: Theory, Models and Applications, Proceedings of an ESCA Workshop. European Speech Communication Association (A. Botinis, G. Kouroupetroglou, & G. Carayannis, editors), pp. 337-340. Athens, Greece: European Speech Communication Association.

WHAT CAN TONE STUDIES TELL US ABOUT INTONATION?

Yi Xu
Department of Communication Sciences and Disorders,
Northwestern University, 2299 North Campus Drive, Evanston, IL 60208: xuyi@nwu.edu

Q. Emily Wang
Department of Communication Disorders and Sciences, Rush University &
Rush-Presbyterian-St. Luke's Medical Center, Chicago, IL 60612: ewang@rush.edu

ABSTRACT

The present paper demonstrates that much can be learned about intonation through the study of the contribution of lexical tones to the f_0 contour of speech utterances. A fundamental principle and several basic mechanisms of tone production and perception are proposed based on studies of both tone and intonation. Their implications for intonation in general are discussed.

1. INTRODUCTION

Whether we assume a linear ([21]) or superpositional ([11]) view of intonation, an unavoidable problem with intonation research has to be resolved sooner or later: the inherent difficulty in determining the *primitives* (basic components) of intonation. For one thing, those primitives cannot be inspected in isolation, because intonation is by definition a global phenomenon. For another, they cannot be manipulated independently of intonation, because they ***are*** part of intonation.

Various efforts have been made in the past to overcome this difficulty. Some use perception to identify the most relevant component of intonation ([28]); others try to divide global intonation curves into local contours, and treat them as linearly concatenated tones ([18], [21]). Neither approach, however, has adequately addressed the problem. Alternative approaches, in our view, should be able to circumvent this dilemma. One such approach, as presented in the present paper, does this by employing a linguistic entity that, though also implemented mainly as f_0 contours, is relatively independent of intonation. This entity is known as lexical tones. The potential effectiveness of this approach lies in the fact that lexical tones can be manipulated independently of intonation. Past studies that look at both tone and intonation, however, have focused mainly on examining how tones are influenced by intonation, but rarely the other way around ([1], [4], [12], [19], [26], [27], [30]). Given the lack of deep understanding of intonation in the first place due to the inherent dilemma, it is not surprising that this kind of investigation still leaves the question of intonation primitives unanswered.

The present study takes a different course — treating lexical tones as contributors to the surface intonation form and using the observed variations of their f_0 contours as clues to the intonation primitives and the basic mechanisms that may underlie both tone and intonation. We will start by laying out the basic mechanisms of tone production and perception, which we have summarized through our own experimentation as well as through extensive review of existing literature on both tone and intonation. We will then proceed to discuss various implications of these mechanisms for our understanding of tone and intonation in general.

2. TONAL MECHANISMS

2.1. The Fundamental Principle

In both production and perception, *a tone is always aligned with the syllable it is associated with*. This principle may sound simplistic given the vast amount of literature seeming to indicate misalignment of the two. However, as we will show, this discrepancy is mainly due to the work of the basic mechanisms of tone production.

2.2. Tone Production Mechanisms

A tone often *appears* to be misaligned with its carrying syllable due to the following mechanisms.

1. A tone is fully realized only in the later portion of a syllable, while the early portion of the f_0 contour in a syllable usually constitutes the transition from the previous tone ([9], [32]) as well as variations due to the voicing characteristic of the initial consonant ([8], [13]). Figure 1 illustrates the work of this mechanism in Mandarin. Note that for any given tone the early portion of the f_0 contour in a syllable always varies with the ending f_0 of the previous syllable, whereas the later portion converges to a contour that conforms to the underlying tone specifications. We call this mechanism *late alignment*.

2. Due to *late alignment*, a dynamic (rising or falling) tone is realized as having a rapid f_0 movement all the way through the end of the syllable. And this rapid f_0 movement often continues into the early portion of the following syllable, presumably due to articulatory inertia, as shown in Figure 2b. We call this mechanism *dynamic overshoot*.

3. The high f_0 region of a tone is usually raised by a following low tone ([9], [17], [32]), as shown in Figure 2. We refer to this mechanism as *anticipatory raising*. Speculations about the nature of this mechanism can be found in Xu & Wang ([33]).

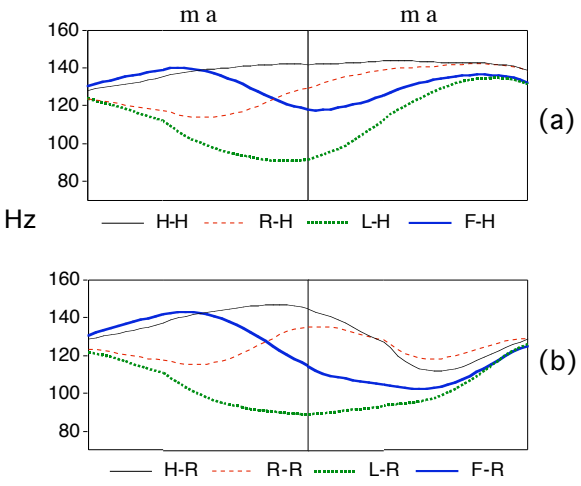

Figure 1. Effects of preceding tone on f_0 contour of following tone in Mandarin. In each panel, the tone of the second syllable is held constant (high-level in panel a, rising in panel b, while the tone of the first syllable is varied: high-level (H), rising (R), low (L), or falling (F). Each curve is an average of 48 tokens produced by 8 speakers. See [32] for details.

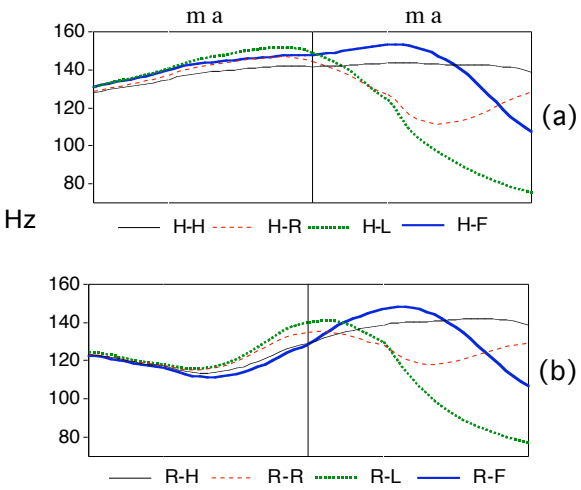

Figure 2. Effects of following tone on f_0 contour of preceding tone in Mandarin. In each panel, the tone of the first syllable is held constant (high-level in panel a, rising in panel b), while the tone of the second syllable is varied: high-level (H), rising (R), low (L), or falling (F). Each curve is an average of 48 tokens produced by 8 speakers. See [32] for details.

2.3. Implications for production and perception of local f_0 contours

The three basic mechanisms described above are responsible for many phenomena of tone production and perception reported in the literature.

2.3.1. Production

First, *late alignment* and *dynamic overshoot* constitute two major aspects of the tonal phenomenon known as *tone spreading,* which has been widely reported in various tone languages ([15]). Tone spreading is described as an assimilation of either a portion of a tone or its entirety to the tone that precedes it ([20]). Thus Figure 1 may be considered as presenting instances of partial tone spreading.

Second, *late alignment* and *anticipatory raising* are probably jointly responsible for the phenomenon known as *downstep* in many tone languages, as illustrated in Figure 3 for Mandarin. Specifically, in a HLH sequence, the second H is lower than the first H because the L both *lowers* the second H and *raises* the first H (Compare HLHHH with HHHHH in Figure 3.).

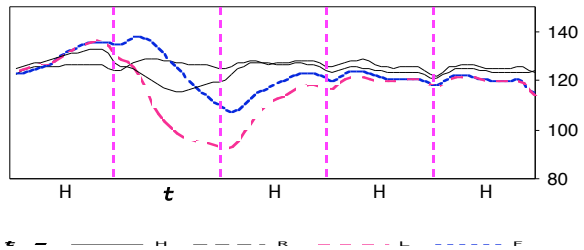

Figure 3. Average f_0 tracings of sentences consisting of all high (H) tones except on the second syllable which carries alternatively high-level (H), rising (R), low (L) and falling (F) tones. Each curve is an average of 20 tokens produced by 4 male speakers (5 repetitions per speaker). See [33] for details.

Third, *dynamic overshoot* is responsible for *peak delay*, the phenomenon that an f_0 peak supposedly associated with a syllable is realized in the following syllable. This phenomenon has been reported for both tone and non-tone languages ([2], [6], [7], [10], [16], [22], [23]).

2.3.2. Perception

In perception, listeners, being also speakers themselves, listen to and hear what the speakers do (or intend to do) when producing tones.

Firstly, they hear only the *later* portion of the f_0 contour in a syllable as the tone proper, while interpreting the *early* portion as carrying information about the previous tone and about the voicing property of the initial consonant ([13], [25]).

Secondly, they listen for whether the tone has a level contour of some value (high, mid, or low) or a dynamic contour of some direction (rising or falling) by focusing on the final portion of the syllable. This pattern of perception is evident not only in the perception of lexical tones ([31]), but also in non-tone languages as suggested by the pitch perception model proposed by House ([14]). According to the House model, there are two classes of basic tone patterns, level tones (high or low) and dynamic tones (rising or falling). This model is surprisingly similar to the basic tone patterns in Mandarin, suggesting that these perception patterns are quite general and are probably rooted in tone production

rather than in perception.

Thirdly, they probably do not hear a delayed f_0 peak as a pitch peak being delayed, but simply as a sharp rising tone, because as also speakers themselves, they "know" that when producing a sharp rising tone, the peak will fall outside of the syllable (due to *dynamic overshoot*). Supporting evidence for this hypothesis may be found in the House model of tone perception [14]. Though claimed to be universal, the models gives no place for the perception of the location of f_0 peaks independent of f_0 contours.

3. INTONATIONAL MECHANISMS

The nature of lexical tones also makes the investigation of more global intonational phenomena easier. We will discuss four global intonation patterns, three of which are directly observable in both tone and non-tone languages, whereas the fourth has been observed only in non-tone languages.

3.1. Declination

There have been two major types of accounts for declination: physiological and non-physiological. The former assume that there are some kind of physiological constraints that force speakers to lower their f_0 over the course of an utterance, while the latter assume that there is some kind of declination line that speakers deliberately follow. The difficulty with the physiological accounts is that it is common knowledge that even musically untrained people can easily sustain a musical note for the duration of several sentences. Why should we then assume that when they speak, they somehow lose the strength to control their pitch long enough and have to let it trail off drastically, sometimes even over just a few syllables (according to many studies)? For the non-physiological accounts, instances such as the neutral-focused sentence shown in Figure 4 (thin solid line) present yet the greatest difficulty. In a sentence consisting of only the high tone, no apparent declination can be observed if there is no particular focus in the sentence. Similar findings were made by other tone studies as well ([17], [27]). We are thus faced with a real dilemma: If declination is not due to physiological constraints, it must have been produced on purpose; yet when all the intended pitch targets are controlled, it is no longer there! The only sensible alternative explanation would then be, that it is not the global trends, but more local shapes, and it is not some obscure declination line, but certain linguistically meaningful intonation melodies, that are deliberately produced, which are ultimately responsible for the surface f_0 contours that can often be described as a global declination. The following few sections will look at some such intonational melodies that we have seen.

3.2. Sentence Focus

The realization of sentence focus, as found in many studies including our own ([5], [33]), can dramatically change the f_0 contour of a sentence. As shown in Figure 4, a sentence consisting of all high-level tones in Mandarin may assume drastically different overall f_0 curves depending on different focus conditions. Briefly, around a sentence focus, there is a radical *asymmetry* in f_0 adjustment: f_0 value *on* the focused word is substantially raised; f_0 value *after* the focus is dramatically lowered; whereas f_0 *before* the focus does not deviate much from the neutral-focus condition. In fact, there can be said to be three distinct focus-related pitch registers, each about 1/2 octave apart, as listed in Table 1.

Compared to lexical tones, focus effect certainly has more global influence on intonation. And it also seems to coexist compatibly with lexical tones.

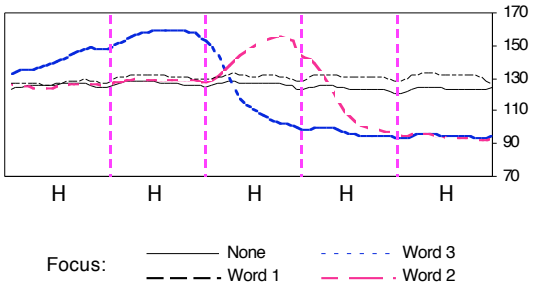

Figure 4. Focus effect in Mandarin. All sentences consist of high-level (H) tones only. Each curve is an average of 20 tokens produced by 4 male speakers.

Table 1. Focus-related pitch registers

Pitch Register	Domain of Assignment
high	non-final on-focus word
low	post-focus words
neutral	all other words

3.3. Introduction of New Ideas or New Topics

A strong case for its effect on intonation is made by Umeda ([29]), who suggests that exceedingly high f_0 peaks at the onset of a sentence are beginning signals for new ideas, or so produced to draw listeners' attention. Cooper, Eady and Mueller ([5]) find that in a moderately long (17 words) sentence in English, the initial f_0 peak is higher than even the f_0 of the focused word that occurs later in the sentence, and there is a sizable f_0 drop across the first and second content words. So the initial f_0 peak seems to be independent of the sentence focus. Cooper et al.'s data indicate that this initial drop accounts for a big chunk of the entire decline in the sentence, while the focus effect accounts for much of the rest.

This kind of non-linear f_0 decline over the course of the sentence fits the exponential downstep model proposed by Liberman & Pierrehumbert ([18]) quite well. Prieto, Shih and Nibert ([24]) also find the exponential model fitting their Spanish data very well. However, similar non-linear f_0 decline is somehow mysteriously missing

in languages that have lexical tones, such as Mandarin ([33]), Yorùbá ([17]), and Swedish ([3]). One possibility is that those languages already use up the manageable pitch ranges for lexical tones and sentence focus, and possibly for other linguistically meaningful pitch melodies, such as final raising for yes/no questions, and there is no more room left for a distinctively high initial pitch target for initiating new topics.

3.4. Downstep

Downstep is often used interchangeably with declination or regarded as the mechanism for it ([18], [24]). Our Mandarin data indicates, however, at least in Mandarin, the scope of downstep caused by a L tone intervening two H tones is rather local ([33]) (cf. Figure 3). Hence, it may not be a major contributor to the more global decline trend that is generally referred to as declination.

4. SUMMARY

To sum up, we have shown that the study of lexical tone can provide us with information on basic tonal and intonational mechanisms that cannot be easily found by looking at intonation alone. In particular, we have demonstrated that the fundamental principle and basic mechanisms we proposed are possible mechanisms for pitch production and perception in both tone and non-tone languages. Furthermore, we have shown that real understanding of more global intonation patterns is contingent upon the comprehension of these basic tone mechanisms.

5. REFERENCES

[1] Abramson, A. S. & Svastikula, K. (1983) Intersections of tone and intonation in Thai. *Status Report on Speech Research, Haskins Laboratories*, **SR-74/75,** 143-167.

[2] Arvaniti, A. & Ladd, D. R. (1995) Tonal alignment and the representation of accentual targets. In *Proceedings of The 13th International Congress of Phonetic Sciences*, Stockholm, **4,** pp. 220-223.

[3] Bruce, G. (1982) Developing the Swedish intonation model. *Lund University, Dept. of Linguistics Working Papers*, **22,** 51-116.

[4] Connell (1983) Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *Journal of Phonetics*, **11,** 337-351.

[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] de Jong, K. & McDonough, J. (1993) Tone and tonogenesis in Navajo. *UCLA Working Papers in Phonetics*, **84**.

[7] de Jong, K. & McDonough, J. (1994) Initial tones and prominence in Seoul Korean. *OSU Working Papers in Linguistics*, **43,** 1-14.

[8] Gandour, J. T. (1974) Consonant types and tone in Siamese. *Journal of Phonetics*, **2,** 337-350.

[9] Gandour, J., Potisuk, S., & Dechongkit, S. (1994) Tonal coarticulation in Thai. *Journal of Phonetics*, **22,** 477-492.

[10] Grimm, C. (1997) *The Phonetic Realization of Pitch Accent in the Ontario Dialect of Oneida*. Unpublished M.A. thesis, State University of New York at Buffalo, Buffalo, New York.

[11] Grønnum, N. (1995) Superposition and subordination in intonation — a non-linear approach. In *Proceedings of The 13th International Congress of Phonetic Sciences*, Stockholm, **2,** pp. 124-131.

[12] Ho, A. T. (1976) Mandarin tones in relation to sentence intonation and grammatical structure. *Journal of Chinese Linguistics*, **4,** 1-13.

[13] Hombert, J.-M. (1978) Consonant types, vowel quality, and tone. In *Tone: A linguistic survey* (V. A. Fromkin, editor), pp. 77-111. New York: Academic Press.

[14] House, D. (1990) *Tonal Perception in Speech*. Lund: Lund University Press.

[15] Hyman, L. & Schuh, R. (1974) Universals of tone rules. *Linguistic Inquiry*, **5,** 81-115.

[16] Ladd, R. D. (1983) Phonological features of intonational peaks. *Language*, **59,** 721-759.

[17] Laniran, Y. (1992) *Intonation in Tone Languages: The phonetic Implementation of Tones in Yorùbá*. Unpublished Ph.D. Dissertation, Cornell University.

[18] Liberman, M. & Pierrehumbert, J. (1984) Intonational invariance under changes in pitch range and length. In *Language Sound Structure* (M. Aronoff & R. Oehrle, editors), pp. 157-233. Cambridge, Massachusetts: M.I.T. Press.

[19] Lindau, M. (1986) Testing a model of intonation in a tone language. *Journal of the Acoustical Society of America*, **80,** 757-764.

[20] Manfredi, V. (1993) Spreading and downstep: Prosodic government in tone languages. In *The Phonology of Tone* (H. van der Hulst & K. Snider, editors), pp. 133-184. New York: Mouton de Gruyter.

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

[22] Pierrehumbert, J. & Steele, S. A. (1989) Categories of tonal alignment in English. *Phonetica*, **46,** 181-196.

[23] Prieto, P., Santen, J. van, & Hirshberg, J. (1995) Tonal alignment patterns in Spanish. *Journal of Phonetics*, **23,** 429-451.

[24] Prieto, P., Shih, C.-L., & Nibert, H. (1996) Pitch downtrend in Spanish. *Journal of Phonetics*, **24,** 445-473.

[25] Rose, P. J. (1988) On the non-equivalence of fundamental frequency and pitch in tonal description. In *Prosodic Analysis and Asian Linguistics: To Honour R. K. Sprigg* (D. Bradley, E. J. A. Henderson, & M. Mazaudon, editors), pp. 55-82. Canberra: Pacific Linguistics.

[26] Shen, X. (1991) Interplay of the four citation tones and intonation in Mandarin Chinese. *Journal of Chinese Linguistics*, **17,** 61-69.

[27] Shih, C.-L. (1988) Tone and intonation in Mandarin. *Working Papers, Cornell Phonetics Laboratory*, **3,** 83-109.

[28] 't Hart, J. & Cohen, A. (1973) Intonation by rule: a perceptual quest. *Journal of Phonetics*, **1,** 309-327.

[29] Umeda, N. (1982) "F_0 declination" is situation dependent. *Journal of Phonetics*, **10,** 279-290.

[30] Vance, T. J. (1976) An experimental investigation of tone and intonation. *Phonetica,* **33,** 368-392.

[31] Whalen, D. H. & Xu, Y. (1992) Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, **49,** 25-47.

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

[33] Xu, Y. & Wang, Q. E. (forthcoming) Intonation contours decomposed — from bottom up.