In *Pragmatics, Prosody and English Language Teaching*. J. Romero-Trillo (ed.). New York: Springer, pp. 61-76, 2012.

Function vs. form in speech prosody — Lessons from experimental research and potential implications for teaching

Yi Xu

Department of Speech, Hearing and Phonetic Sciences

Division of Psychology and Language Sciences

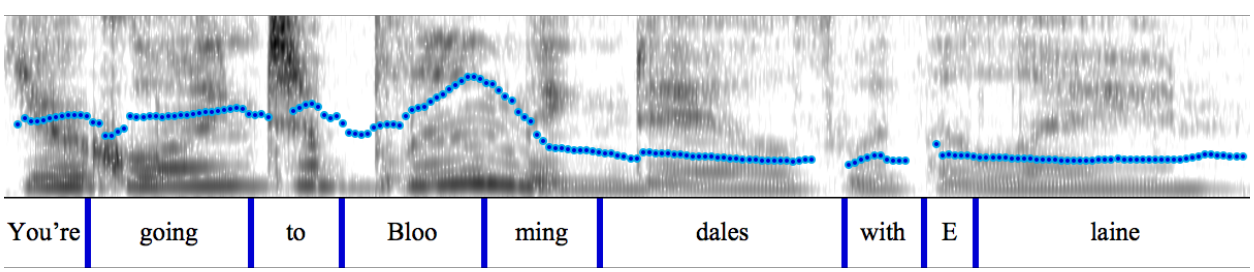
University College London, UK

Key words: articulatory mechanism, target approximation, pitch range, communicative function, tone, focus, sentence type

1. INTRODUCTION

Prosody is an indispensable part of a language and so its mastery is highly desirable for learners of any language, especially those at a relatively advanced level. But the pressing question is how this can be achieved, as prosody is notoriously difficult to learn (Atoye, 2005; Dankovicova et al., 2007), and so far there has been a lack of effective ways to teach it (Atoye, 2005). This difficulty is closely linked to the fundamental question of what is it that needs to be learned in terms of the prosody of a language. To start with an example, Figure 1 displays the pitch tracts (dotted lines) of two sentences in General American English spoken by a female native speaker. The top one is a statement and the bottom one a question. In both sentences, the speaker puts emphasis on "Bloomingdales", which means that the word is the location of the nucleus (British) or nuclear accent (Pierrehumbert, 1980). Whatever following the large pitch movement should then be considered as the tail. As for what kind of nuclear it is, the top one could be considered as rise-fall, and the bottom one low rise. But the question is, are these the patterns that learner of English need to be taught?

You're going to ***Bloomingdales*** with Elaine.



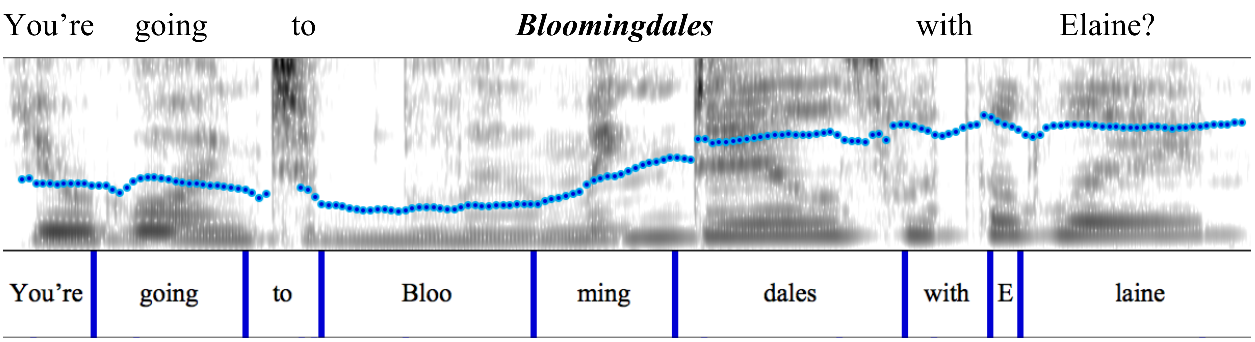


Figure 1. Spectrograms and F_0 tracks of the sentence "You're going to Bloomingdales with Elaine", spoken by a female American English speaker as either a statement (a) or a question (b). Samples from Liu & Xu (2007).

To answer this kind of questions, experimental findings on tone and intonation will first be reviewed, and then potential implications of these findings for the teaching of English as a second language will be explored.

2. BASIC ARTICULATORY MECHANISM — LESSONS FROM TONE LANGUAGES

Figure 2a displays mean F_0 contours of four Mandarin lexical tones said in isolation, averaged across 40 repetitions by eight male speakers. These contours can be considered as close to the canonical forms of these tones, i.e., the underlying patterns, because they are free of influence from surrounding tones. In Figure 3, the same four tones are produced in a five-syllable sentence by 4 males speakers. In the second syllable of each graph, we can see that the contours of the four tones bear much resemblance to those in Figure 2a, with the exception of the L tone whose final rise in Figure 2a is missing in Figure 3, presumably due to a phonological rule, as will be discussed later (Chao, 1968). Syllable 3 in Figure 3, however, shows rather different F_0 contours despite the fact that its tone remains constant: H, R and F in (a), (b) and (c), respectively. These variations appear to be directly related to the fact that the F_0 contours of the four preceding tones have very different endings, and the beginning portion of the tone in syllables 3 seems to be direct continuations of those endings. Interestingly, however, by the end of syllable 3, the four contours have virtually converged to a straight line whose height and slope are consistent with canonical forms of the tones: high-level for H, rising for R and falling for F. Similar patterns of contextual F_0 variations have been found in other tone languages (Gandour et al. 1994 for Thai, Wong, 2006 for Cantonese, and Laniran & Clements, 2003 for Yoruba). From these studies we can see the following mechanisms of tone production:

1) A tone is produced by approaching its underlying canonical target *within* the syllable that carries it, i.e., from the syllable onset to the offset. This means that the tone is articulated in synchrony with its host syllable. Such synchrony is maintained regardless of whether the initial consonant of the syllable is voiced or whether the syllable has a coda consonant (Wong & Xu, 2007 for Cantonese; Xu & Xu, 2003 and Xu, 1998 for Mandarin).

2) It takes a substantial amount of time for F_0 to go from the ending value of the preceding tone to the desired value of the current tone. This is because, according to Sundberg

(1979) and Xu and Sun (2002), at least a 100 ms is needed to make a pitch change of even the smallest size, and the amount of time further increases with the size of the pitch change. Based on data from Xu (1997, 1999), a greater half of a syllable of an average duration of 200 ms is needed to make most of the tonal transitions. Furthermore, for the F tone in syllable 3 in Figure 3a, because two pitch movements are involved within a single syllable, there is varying degrees of undershoot of the tonal target depending on the tone of syllable 2.

3) As a consequence of these basic tone production mechanisms, F_0 movements in the early half of the syllable mostly serve as transitions to the underlying tonal targets. And as such they cannot be taken as the underlying tonal contours per se. The F_0 contours toward the end of the syllable, in contrast, seem to best resemble the underlying tonal targets.

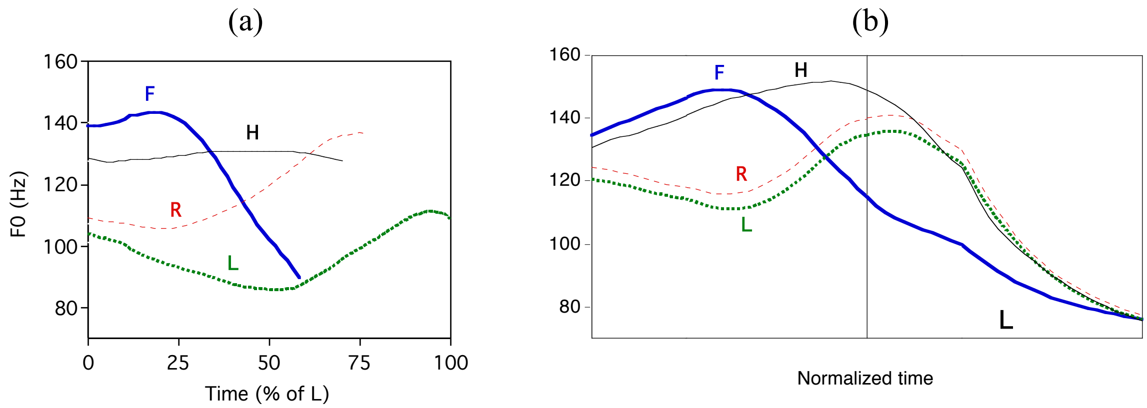


Figure 2. (a): Four Mandarin tones produced in isolation. (b): Mandarin L tone after four different tones, produced in carrier phrases. Adapted from Xu (1997).

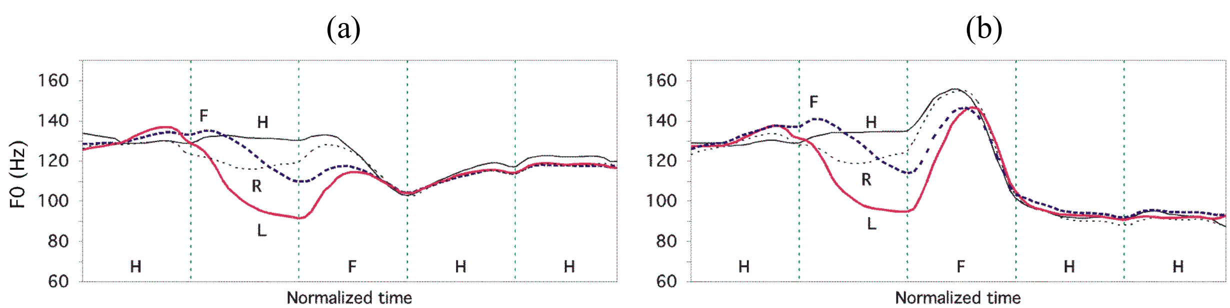


Figure 3. Mandarin tone F following four different tones. (a): no narrow focus in the sentence; (b): focus on the F-carrying syllable. Each curve is an average of 20 tokens produced by four male speakers (five repetitions per speaker). (Data from Xu 1999). √

The basic mechanism of tone articulation has been summarized into the Target Approximation (TA) model, as illustrated in Figure 4 (Xu & Wang, 2001). According to the model, lexical tones are produced by a process of articulatorily approaching successive local pitch targets, each in synchrony with the host syllable. In this model, there are several parameters that can be controlled by the speaker: target height, target slope and target strength (which defines the speed at which the target is approached), pitch range (which defines the overall height and vertical span of the target). The approximation of the target is always synchronized with the entire syllable, i.e., commencing with the syllable onset and terminating with the syllable offset (regardless of whether the target has been reached). Thus the synchronization of pitch target and the syllable is assumed to be obligatory, leaving no room for the speaker to adjust the timing of the target relative to the temporal interval of the syllable. Finally, the TA model assumes that the final F_0 state of a syllable is transferred across the syllable boundary to become the initial state of the next syllable, as illustrated in Figure 4.

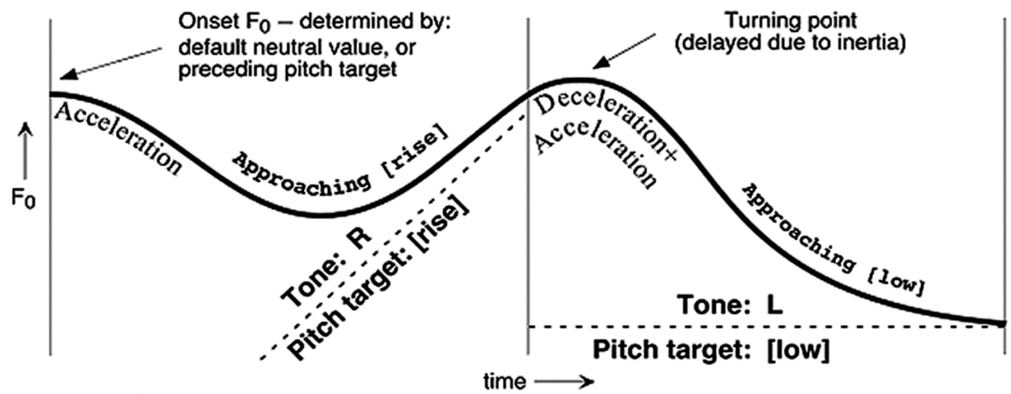


Figure 4: Target approximation model, adapted from Xu and Wang (2001).

3. RELEVANCE FOR NON-TONE LANGUAGES

One could argue that the articulatory mechanism of tone production is applicable only to tone languages, and they are necessitated by the fact that in a language like Mandarin or Cantonese, every syllable is specified with a lexical tone and therefore it is articulatorily more demanding than a non-tone language like English (Ladd, 1996). There is evidence, however, that these mechanisms also apply to non-tone languages. The investigation by Xu and Sun (2002) examined native speakers of both Mandarin and English, and virtually no difference was found between the speakers of the two languages in terms of maximum speed of pitch change. Thus native Mandarin speakers do not change pitch faster than native English speakers despite years of speaking a tone language. In the same study, previously reported data from Dutch and English were also reviewed and it was concluded that speakers of those two languages often approach the maximum speed of pitch change in their own speech (Caspers & van Heuven, 1993). There is also evidence, either direct or indirect, that pitch production is synchronized with the syllable in English and other non-tone languages. Xu and Wallace (2004) show that F_0 transitions similar to those in Figure 3 start from the syllable onset in American English regardless of whether the initial consonant is voiced or voiceless. Similar evidence was reported earlier by Silverman (1986). Further evidence of syllable-synchronization of pitch production as well as perception has been shown by Gao and Xu (2010) for Southern British English, Dilley (2007) for American English and Niebuhr (2007) for German. The existing evidence therefore demonstrates that native English speakers change pitch no slower than speakers of a tone language, and they also likely produce underlying pitch targets in synchrony with the syllable.

Assuming that mechanisms of producing pitch events in English is similar to those of tone languages, interpretation of F_0 contours in English should also follow that of tonal contours. In Figure 1a, for example, the F_0 rise in the first syllable of "Bloomingdales" appears very sharp. But because the rising movement starts from the syllable onset and ends before the syllable offset, it should not be considered as due to an underlying rising target. Rather, it should be more like the Mandarin H tone shown in Figure 5a, which means that the underlying target of the stressed syllable in "Bloomingdales" is likely to be a static high at least in this particular example. In contrast, in Figure 5b, we can see that the R tone in Mandarin, which presumably does have a genuine rising target, exhibits a clear F_0 rise only in the second half of the syllable, and this rise continues into the beginning part of the syllable. This "peak delay" is likely due to articulatory inertia, and is modeled by the TA model as resulting from transferring the final state of the previous F_0 movement to the beginning of the next syllable, as shown in Figure 4. Similarly, the F tone, which presumably has a falling target, exhibits a clear fall in the second half of the syllable, as shown in Figure 5c. Finally, the F_0 in the second syllable of "Bloomingdales" shows a sharp fall. However, because the F_0 movement is virtually leveled off before the syllable offset, the underlying pitch target is unlikely to be a [fall], but rather more likely to be a relatively low static register, similar to the Mandarin H tone in syllable 4 in Figure 5a, whose height is substantially lowered by focus, as will be discussed later. More importantly, the sharp fall right after the F_0 peak, because it mostly occurs in the early part of the unstressed syllable after "Bloo-", should not be considered as due to a falling target.

4. FUNCTION VERSUS FORM

To speak is to convey meanings, and prosody is an important meaning carrier. In the traditional approaches, however, prosodic categories are primarily defined by their forms, while their functional connotations are usually left vague. In the nuclear tone tradition, the forms of various types of nucleus are clearly defined: Rise, Fall, Rise-fall, Fall-rise, High fall, Low fall, High rise, Low rise, etc. (Palmer, 1922; Crystal, 1969; Halliday, 1967; O'Connor and Arnold, 1961). Bolinger, who has put much emphasis on the importance of communicative values of prosody, also first defines the basic intonational components in terms of their forms, e.g., A accent, B accent, etc. (Bolinger, 1986). In the Pierrehumbert model, from which the Tonal elements of the ToBI transcription system were derived, intonational components are deemed to be phonological (i.e., contrastive) without establishing their meanings. In her own words,

> "In the literature, one can distinguish two approaches towards the problem of establishing which intonation patterns are linguistically distinct and which count as variants of the same pattern. One approach attacks the problem by attempting to deduce a system of phonological representation for intonation from observed features of F0 contours. After constructing such a system, the next step is to compare the usage of F0 patterns which are phonologically distinct. The contrasting approach is to begin by identifying intonation patterns which seem to convey the same or different nuances. The second step is to construct a phonology which gives the same underlying representation to contours with the same meaning, and different representations to contours with different meanings… *The*

> *work presented here takes the first approach, in fact, it stops at the first step in the first approach* [my emphasis].” (p. 59)

It is not the case that traditional approaches are unconcerned with functions. Rather, they typically treat the forms as the defining properties of intonational categories. Functions, in contrast, are viewed as simply accompany the forms. Pierrehumbert and Hirschberg (1990), for example, made an attempt to identify the meanings associated with the form-based intonational units established in Pierrehumbert (1980). The meanings identified in the paper, however, are long-winded, vague and heavily overlapped with each other. In addition, probably more critically, tonal units such as H and L are conceptualized as directly meaningful. For example, the following are descriptions of the meanings of the H* and L* pitch accents:

> The H* accents above and in utterances in general convey that the items made salient by the H* are to be treated as "new" in the discourse. (p. 289)

> The L* accent marks items that S intends to be salient but not to form part of what S is predicating in the utterance. (p. 291)

Note that the two aspects of the traditional approaches, namely, a) giving form priority and b) assigning meanings directly to intonational units, are rather different from how the meaning-form relation is viewed in the case of the segmental aspect of speech. A segmental contrast is defined, first and foremost, by whether it can distinguish words or grammatical functions. Those differences that do not distinguish one word from another, e.g., the many /r/ variants in English, are considered as allophones rather than separate phonemes. Therefore, lexical contrast, as a function, is the *defining property* rather than a *subordinate* or *accompanying property*. Secondly, lexical contrast itself does not carry any specific meanings, but only serves to distinguish lexical items from one another. The specific meanings of the words are defined in morphology rather than in phonology, and only occasionally, e.g. in the case of onomatopoeia, is there any direct link between phonetic form and lexical meaning. Similarly, in the lexical use of pitch by either lexical tone or lexical stress, again pitch itself carries no specific meaning, but only serves to distinguish words. Nevertheless, there is a further aspect that is more unique to tone and intonation, but less of a concern with segments. That is, segmental contrasts serves to distinguish only lexical items. But tonal and intonational components carry multiple functions. The traditional approaches, however, usually do not try to keep the functions clearly separated from each other. The meaning definitions for the H* and L* pitch accents cited above exemplify the problem. In both cases, salience, newness and predication should be separate functions. But they are put together as components of a single hypothetical prosodic unit, because the F_0 shape is treated as the defining property.

An important reason for the persisting practice of giving primacy to prosodic form is that identifying functional categories in prosody from acoustic signals is difficult. This is because there are at least three degrees of separation between surface prosodic forms and the communicative functions they encode (Xu, 2004). The first degree of separation is due to the articulatory mechanisms discussed in the previous section. Because of these mechanisms, directly observed surface acoustic forms often do not resemble the relatively invariant underlying phonetic targets that are used to encode information. It is only through systematically controlled experiments can the invariance of the underlying targets be observed. The second degree of separation is due to target reassignment, which is a process of changing the underlying targets depending on various factors such as phonetic context and prosodic functions. A case in point is the Mandarin L tone, which has a rising tail when produced in isolation or sometimes in a sentence-final position. But the tail is missing when the L is followed by any other tone, as can be seen in Figure 3, and its absence cannot be explained by articulatory mechanisms. In addition, the L tone in Mandarin changes into a R tone when followed by another L tone. This change again has no plausible articulatory explanations. Such target reassignment happens not only to lexical tones, but also to lexical stress as a function of other prosodic functions, as will be discussed subsequently. The third degree of separation is due to parallel encoding of multiple communicative functions, as will be elaborated next.

4.1. Lexical versus extra-lexical functions

Because F_0 is the major acoustic correlate of lexical tones, it could be assumed that intonational use of pitch is very limited in a tone language. But this assumption can be easily dismissed by some basic facts about pitch perception and production. On the one hand, native listeners can identify the four Mandarin tones in their native language with a pitch differences no larger than 0.5 semitones (or 4 Hz) (Klatt, 1973). On the other hand, the natural pitch range of an average speaker is well over two octaves, i.e., 36 semitones (Honorof & Whalen, 2005). Thus there is plenty of room for the use of pitch to encode both tone and intonation in a tone language. This is exactly what is found in languages like Mandarin. Figure 5a displays time-normalized mean F_0 contours of four five-syllable sentences in Mandarin, which differ from each other only in terms of the tone of the second syllable. Clear differences due to the tone of the second syllable can be seen. In Figure 5b, the same sentences are spoken with prosodic focus on the first word consisting of two syllables. The differences in F_0 due to the tones of the second syllable are clearly exaggerated: The high pitch becomes even higher and the low pitch even lower. The differences between the sentences in Figure 5a and 6b therefore constitute coding for the prosodic focus. Thus both tone and focus are effectively encoded in parallel in Mandarin.

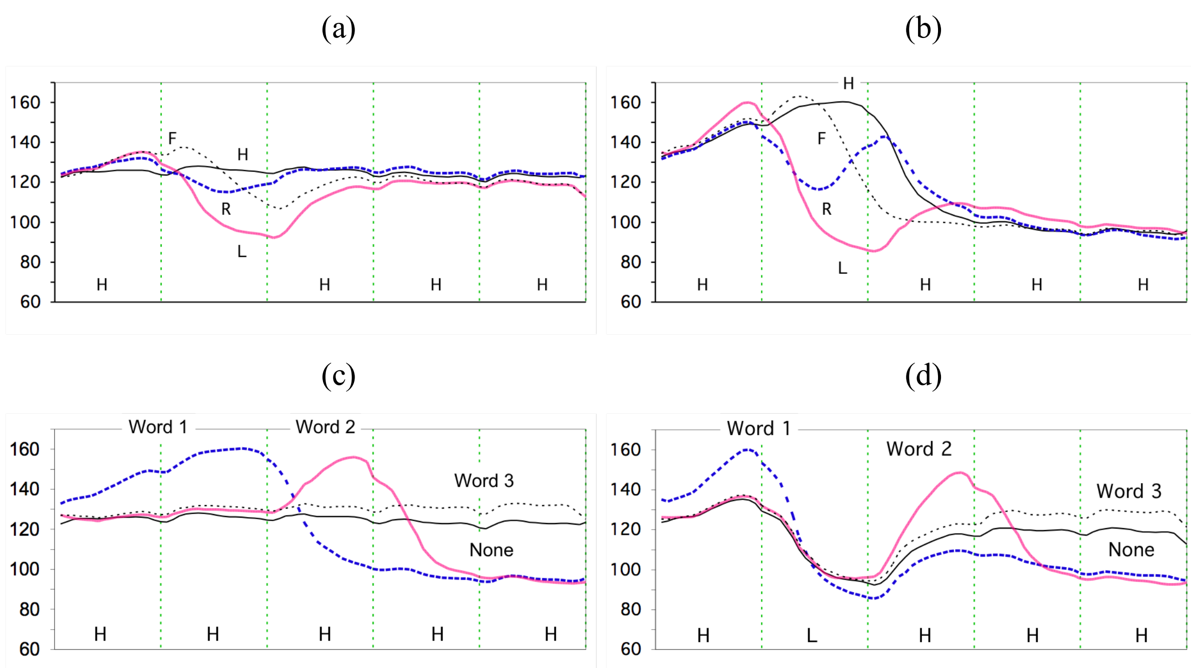
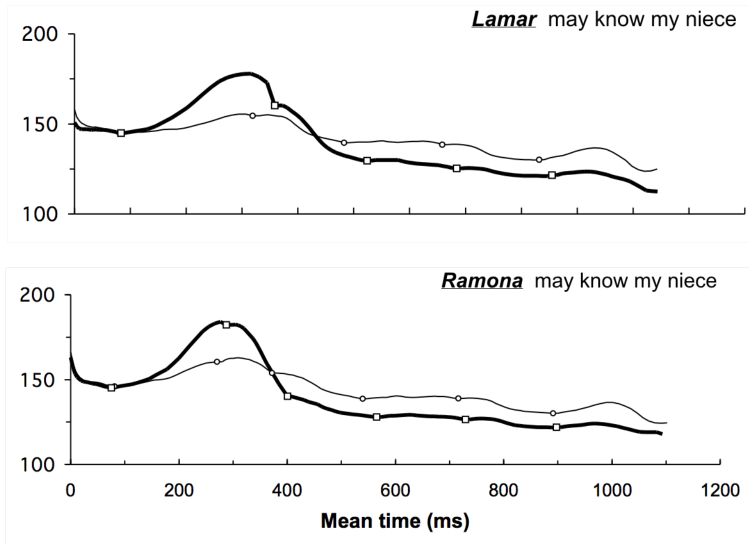


Figure 5. (a-b): Time-normalized mean F_0 contours (20 tokens by four male speakers) of four five-syllable sentences in Mandarin, which differ from each other in terms of the tone of the second syllable. In (a) there is no narrow focus in the sentence; in (b) the first two syllables are in focus. (c-d): Two sentences with tone sequences of HHHH and HLHHH with focus on word 1 (syllables 1-2), word 2 (syllable 3) or word 3 (syllables 4-5). (Data from Xu 1999).

In Figure 5b we can also see that after the exaggerated F_0 contours of the first two syllables, the pitch level of all the subsequent syllables are lowered relative to that in Figure 5a. This can be more clearly seen in Figure 5c and 5d, in which focus varies from word 1, word 2, word 3 and none. It is obvious that as long as focus is not sentence-final, there is significant lowering and narrowing of post-focus pitch range. Chen et al. (2009) also find that post-focus intensity is also significantly reduced. The reduction of pitch range and intensity after prosodic focus is known as *post-focus compression* (Chen et al., 2009), or PFC for short (Xu et al., in press).

Parallel encoding of lexical and extra-lexical functions has also been found in English, a non-tone language. Fry (1958) shows that listeners use the pitch difference between two adjacent syllables to determine which is lexically stressed. The syllable with the higher pitch is heard as stressed. He finds that a difference as small as 5 Hz leads to unambiguous judgment of lexical stress. What Fry has found is only about how lexical stress is used to distinguish words in English, because listeners in his experiments were asked to judge whether a word like "digest" or "permit" is a noun or a verb, but not whether any syllable is stressed. Thus a five Hz difference in F_0 is sufficient to indicate the functional contrast of lexical distinction. Such a functional use of F_0 is actually very similar to lexical tone, and the similarity is especially high in comparison with the neutral tone in Mandarin (Chen & Xu, 2006). Also similar to Mandarin, the focus function is overlaid on the lexical function (Cooper et al., 1985; Pell, 2001; Xu & Xu, 2005). Figure 6 shows mean F_0 contours of two English sentences spoken by seven native American English speakers. The thin curves are from sentences said without prosodic focus, in which the small F_0 undulations correspond to the relative stress of the individual syllables. When a word is focused, as shown by the thick curves, the F_0 contour of the focused word is increased, but the increase does not seem to create any new peaks in addition to those already in the neutral-focus curve. Thus focus seems to expands the pitch range of the focused words relative to that of the neutral focus F_0. Furthermore, the F_0 of the post-focus words is lowered, but again the lowering does not seem to eliminate F_0 peaks from the neutral-focus F_0 curves. Xu and Xu (2005) found that the height of these small F_0 "bumps" is actually comparable to the minimal F_0 difference needed for the perception of lexical stress according to Fry's (1958) study. So the pitch range of post-focus words is reduced, just like in Mandarin. Therefore we can see that lexical and focal contrasts are both encoded with F_0, and the encoding of the two are done in parallel, allowing both to be sufficiently distinctive.



(a)

(b)

Figure 6. Time-normalized mean F_0 contours (each over 49 tokens by seven speakers) of two sentences in American English. The thin curves are from sentences said without prosodic focus; the thick curves are from sentences with focus on the first word. Adapted from Xu & Xu (2005).

More interestingly, again like Mandarin and many other Chinese languages, target reassignment also occurs in English, although the conditions of its occurrence is different. One of the conditions that trigger target reassignment is focus. In Figure 6a the final portion of the F_0 contour levels off toward the end of the stressed syllable /mar/ in "Lamar" when the sentence has no narrow focus (thin line). When "Lamar" is focused, however, the final portion of /mar/ becomes falling. The fall is not as sharp as that of the F tone in Mandarin shown in Figure 5, but sharper than that of the Mandarin High tone shown in Figure 5. More detailed examinations show that individual English speakers differ in terms of how sharp the fall is. But overall, there seems to be a tendency to change the targeted F_0 trajectory of a word-final stressed syllable from [high] to [fall]. This focus-triggered target change is highly condition-specific, because it does not happen to stressed syllables that are not word final. In Figure 6b, for example, F_0 continues to rise toward the end of the stressed syllable /mo/ in "Ramona" whether or not the word is in focus.

Thus in both Mandarin and English, F_0 is used to simultaneously encode lexical and focal contrasts. The two are encoded in parallel by employing different encoding strategies. Lexical contrast is encoded mainly with syllable-sized pitch targets, while focal contrast is encoded mainly by modifying the pitch range of the local targets. For English, however, focus also interacts with lexical stress by changing the local targets of word-final stressed syllables.

4.2. Encoding of sentence type together with lexical stress and focus

Beside lexical stress and focus, another function that further shapes English prosody is sentence type, which determines whether an utterance is spoken as a statement or a question. This is achieved again through parallel encoding and target reassignment (Eady & Cooper, 1986; Liu & Xu, 2007), as illustrated in Figure 7. The word "job" in Figure 7a and 7b and the stressed final syllable of "massage" in Figure 7b both show a sharp F_0 fall toward the end of the syllable when they are under focus in a statement (dashed line), which is consistent with what was described in the previous section. But when the sentences are questions, these syllables all show a sharp F_0 rise all the way to the end of the syllable. This indicates that the underlying pitch target of the syllable, which is stressed and word-final, changes from falling to rising. In addition, Liu and Xu (2007) find that the pitch targets of non-word-final stressed syllables also tend to have a rising target, although the slope of the rise is not nearly as harp as that of word-final stressed syllables. This can be seen in the first syllable of "Microsoft" in Figure 7a.

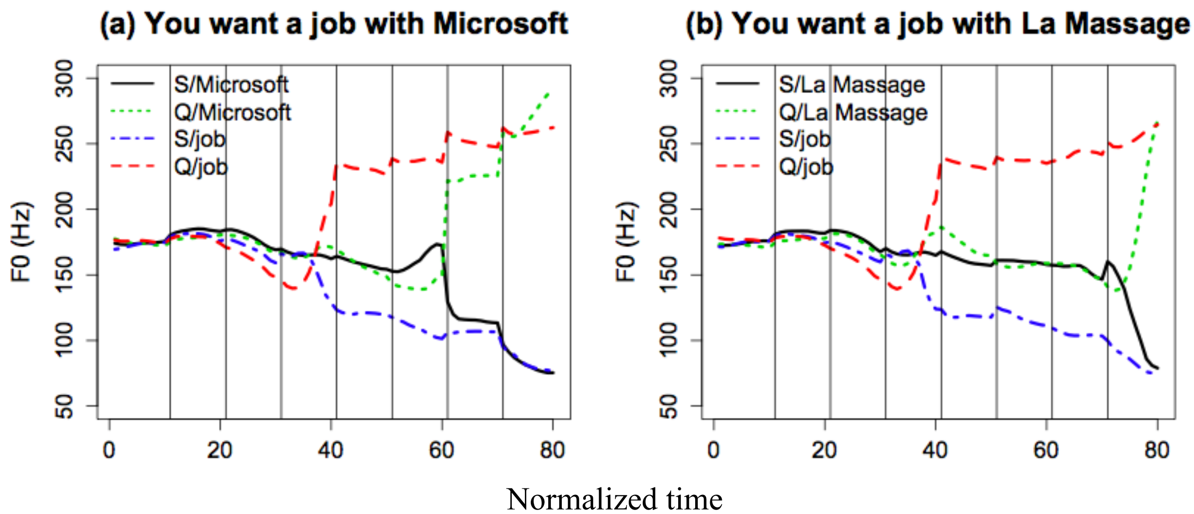


Figure 7. Mean F_0 contours of American English statements and questions with focus. The word after "/" is focused. S = statement, Q = question. The vertical lines indicate syllable boundaries. Data from Liu & Xu (2007).

In addition to the local F_0 excursions, Figure 7 also shows that the height of the F_0 contour is dramatically increased immediately after the focused stressed syllable in a question, and it continues to rise all the way toward the end of the sentence. This pattern contrasts with the dramatic post-focus drop of F_0 height in a statement. Meanwhile, the post-focus local pitch movements are very small in both statements and questions. It appears that post-focus compression of pitch range in terms of local excursions applies in English whether the sentence is a statement or question, but the overall height is very different, high in a question but low in a statement. In other words, the overall height of post-focus pitch range is functionally determined by sentence type, whereas the magnitude of the local excursions is closely related to focus.

To summarize, much of the variations in English result from the parallel encoding of three basic communicative functions: lexical stress, prosodic focus and sentence type, with the following rules:

1. Each syllable, whether stressed or unstressed, is assigned a local pitch target; the properties of the target is jointly determined by lexical stress, focus and sentence type;
2. In a statement a stressed syllable is assigned a high target unless it is word final and on-focus, in which case it is assigned a falling target;
3. In a question, a stressed syllables is assigned a rising target;

4. Prosodic focus expands the pitch range of the focused word, and compresses the pitch range of all post-focus words, but leaves the pitch range of pre-focus words unchanged from that of the neutral focus sentence;
5. A statement gradually lowers the pitch range throughout the sentence, but adding an extra drop immediately after the stressed syllable on focus and at the end of the sentence;
6. A question does the opposite to the pitch range, i.e., raising it throughout the sentence, but adding an extra upward boost immediately after the focused stressed syllable as well as at the end of the sentence.
7. Unstressed syllables are not targetless, but are assigned a mid target with weak articulatory strength (Xu & Xu, 2005).

4.3. Additional functions

Beside lexical stress, focus and sentence type, there are other communicative functions that are also encoded mainly through prosody. So far, however, there has not yet been empirical research that has generated highly specific descriptions of their patterns. One of the functions is topic or turn initiation, which raises the pitch range of the beginning of a sentence (Lehiste, 1975). Another is known as the contradiction contour, which exhibits a global fall-rise pattern across the whole sentence (Liberman & Sag, 1974). The exact condition of its occurrence is not yet fully clear, however, making it difficult to be experimentally investigated. There are other possible stylistic global patterns among those described in the descriptive literature (Bolinger, 1986, 1989; Cruttenden, 1997; Crystal, 1969; Halliday, 1967; O'Connor and Arnold, 1961). But again the exact function and condition of occurrence of those patterns need to be systematically investigated. Much more empirical research is therefore needed.

5. IMPLICATIONS FOR TEACHING ENGLISH INTONATION

The empirical findings on tone and intonation discussed so far may suggest alternative approaches to the teaching of English intonation that are different from the current practices. The dominant strategy in teaching English intonation, to my knowledge, is based the nuclear tone tradition. As discussed earlier, this tradition is mostly form-oriented. The fact that its effectiveness has yet to be shown (Atoye, 2005; Currie, 1980) demonstrates that new methodology could be considered. Since I am not aware of any empirical research on teaching English intonation based on the new findings, I could only offer some preliminary suggestions.

The first suggestion is that it could be beneficial to teach learners of English *functionally defined* intonation patterns. For example, the summary of how lexical, focal and sentential functions are conveyed through pitch as discussed in previous sections could be developed into possible teaching instructions:

1. In a statement, a stressed syllable should have higher pitch than the an unstressed syllable; its pitch contour is preferably level, unless it is word-final and focused or sentence-final, in which case the contour is preferably falling.

2. In a question, a stressed syllable should have *lower* pitch than an unstressed syllable; its pitch contour is preferably slightly rising, unless it is word-final and focused or sentence-final, in which case the contour should be sharply rising.
3. If a word is focused, the pitch of its stressed syllable should be exaggerated, i.e., becoming higher in a statement but lower in a question;
4. Immediately after the stressed syllable of a focused word, the pitch of all the following syllables in the sentence should be extensively lowered in a statement, but raised in a question.

There are foreseeable difficulties in implementing these suggestions, however. The first is that learners may vary extensively in terms of their ability to follow instructions on both local pitch targets and global pitch patterns, as is found by Dankovicova et al. (2007). As a result, many learners may not be able to follow instructions about pitch patterns. The second potential difficulty is the interference of learners' first language. In particular, recent research has found that there is a typological divide among the world languages in terms of the application of PFC, i.e., the extensive reduction of pitch range after focus, as described in 4.1 and mentioned again in suggestion 4 above. Languages in which PFC applies include Indo-European languages like English, German, Italian, Swedish, etc., Altaic languages like Turkish, Japanese and Korean, Uralic languages like Finnish, Semitic languages like Arabic, and Northern Chinese languages like Mandarin (see Xu, 2011 for a brief summary). Languages in which PFC does not apply include Southern Chinese languages like Cantonese and Taiwanese (Chen et al., 2009; Wu & Chung, 2011), Mon-Khmer languages like Wa, Deang (Wang et al., 2011) and possibly Vietnamese (Jannedy, 2007), and many African languages, including Sotho, Buli, Chichewa and Hausa, as summarized by Zerbian et al. (2010). Native speakers of these, and probably many more "non-PFC" languages thus may have difficulty learning PFC, at least without explicit instructions, as indicated by the findings of a number of recent studies (Chen et al., 2009; Wu & Chung, 2011).

Note, however, these difficulties would occur even with traditional approaches to intonation teaching. A function-based approach, nevertheless, may potentially make it easier for learners to become aware of the most critical components of intonation that are functionally relevant. In any case, the strategy of teaching the interactive patterns of lexical stress, focus and sentence type has never been tested before, to my knowledge. Given the clear findings in the first language research, it is at least worth exploring.

A further suggestion that could be offered is in regard to the articulatory mechanisms of tone production discussed near the beginning of the chapter. That is, given that the obligatory synchronization of local pitch targets with the syllable, there is no need to teach learners the alignment of F_0 peaks and valleys. Such alignment has been much discussed in recent literature based on the framework of the Autosegmental-Metrical Phonology of intonation (Pierrehumbert, 1980), also known as Intonational Phonology (Ladd, 1996). According to the Target Approximation model shown in Figure 4, the exact location of F_0 turning points is a direct consequence of the underlying pitch target of the current syllable and those of the adjacent syllables. If such alignment of the underlying pitch targets is obligatory, there is little room for learning, because learning is possible only when the learners have free choices. The real choices learners have are likely in terms of the properties of the pitch target in terms of height and slope, as shown in Figure 4, but not in terms of its micro-timing within the syllable.

6. CONCLUDING REMARKS

This chapter has provided a brief overview of the latest findings about tone and intonation from an articulatory-functional perspective. It is shown in particular that much of the English intonation can be understood in terms of parallel encoding of multiple communicative functions, including lexical stress, focus and sentence type. Based on these findings, some suggestions are offered in terms of possible alternative teaching strategies that depart from the current common practice. In general terms, first, it might be more effective to teach functionally defined prosodic patterns rather than patterns classified in terms of their surface form. Second, it might be more effective to teach syllable-based pitch targets rather than word- or phrase-based whole-contours or their alignments. Finally, it might be more effective to teach complex prosodic forms as resulting from interactions of multiple communicative functions each with a relatively simple underlying form rather than as prosodic gestalts each with a convoluted set of meaning attributes. Whether these suggestions will lead to improvement in effective teaching of English intonation, however, awaits empirical research that puts them to real test.

7. REFERENCES

Atoye, R. O. (2005). Non-native perception and interpretation of English intonation. *Nordic Journal of African Studies* 14: 26-42.

Bolinger, D. (1986). *Intonation and its parts: melody in spoken English*. Palo Alto: Stanford University Press.

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

Caspers, J. and van Heuven, V. J. (1993). Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica* 50: 161-171.

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

Chen, S.-w., Wang, B. and Xu, Y. (2009). Closely related languages, different ways of realizing focus. In *Proceedings of Interspeech 2009*, Brighton, UK: 1007-1010.

Chen, Y. and Xu, Y. (2006). Production of weak elements in speech -- Evidence from f0 patterns of neutral tone in standard Chinese. *Phonetica* 63: 47-75.

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

Currie, K. L. (1980). An Initial Search for Tonics. *Language and Speech* 23: 329-350.

Cruttenden, A. (1997). *Intonation*. Cambridge: Cambridge University Press.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. London: Cambridge University Press.

Dankovicova, J., House, J., Crooks, A. and Jones, K. (2007). The Relationship between Musical Skills, Music Training, and Intonation Analysis Skills. *Language & Speech* 50: 177-225.

Dilley, L. C. and Brown, M. (2007). Effects of pitch range variation on f0 extrema in an imitation task. *Journal of Phonetics* 35(4): 523-551.

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

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

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

Gao, H. and Xu, Y. (2010). Ambisyllabicity in English: How real is it? In *Proceedings of The 9th Phonetics Conference of China (PCC2010)*, Tianjin

Halliday, M. A. K. (1967). *Intonation and Grammar in British English*. The Hague: Mouton.

Honorof, D. N. and Whalen, D. H. (2005). Perception of pitch location within a speaker's F0 range. *Journal of the Acoustical Society of America* 117: 2193-2200.

Jannedy, S. (2007). Prosodic Focus in Vietnamese. In *Interdisciplinary Studies on Information Structure (ISIS)*. S. Ishihara, S. Jannedy and A. Schwarz: Potsdam University Press pp. 209-230.

Klatt, D. H. (1973). Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception. *Journal of the Acoustical Society of America* 53: 8-16.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

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

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

Liberman, M. and Sag, I. (1974). Prosodic form and discourse function. In *Proceedings of Papers from the 10th Regional Meeting*. Chicago Linguistic Society: 416-427.

Liu, F. and Xu, Y. (2007). Question intonation as affected by word stress and focus in English. In *Proceedings of The 16th International Congress of Phonetic Sciences*, Saarbrücken: 1189-1192.

Niebuhr, O. (2007). The Signalling of German Rising-Falling Intonation Categories –

The Interplay of Synchronization, Shape, and Height. *Phonetica* **64**: 174-193.
O'Connor, J. D. and Arnold, G. F. (1973). *Intonation of Colloquial English*. Second Edition. London: Longmans.
Palmer, H. E. (1922). *English Intonation, with Systematic Exercises*. Cambridge: Heffer.
Pell, M. D. (2001). Influence of emotion and focus on prosody in matched statements and questions. *Journal of the Acoustical Society of America* **109**: 1668-1680.
Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation, MIT, Cambridge, MA. [Published in 1987 by Indiana University Linguistics Club, Bloomington].
Pierrehumbert, J. and Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In *Intentions in Communication*. P. R. Cohen, J. Morgan and M. E. Pollack. Cambridge, Massachusetts: MIT Press pp. 271-311.
Silverman, K. (1986). F0 segmental cues depend on intonation: The case of the rise after voiced stops. *Phonetica* **43**: 76-91.
Sundberg, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics* **7**: 71-79.
Wang, B., Wang, L. and Kadir, T. (2011). Prosodic encoding of focus in six languages in China. In *Proceedings of The 17th International Congress of Phonetic Sciences*, Hong Kong, pp. 144-147.
Wong, Y. W. (2006). Contextual Tonal Variations and Pitch Targets in Cantonese. In *Proceedings of Speech Prosody 2006*, Dresden, Germany: PS3-13-199.
Wong, Y. W. and Xu, Y. (2007). Consonantal perturbation of f0 contours of Cantonese tones. In *Proceedings of The 16th International Congress of Phonetic Sciences*, Saarbrucken: 1293-1296.
Wu, W. L. and Chung, L. (2011). Post-focus compression in English-Cantonese bilingual speakers. In *Proceedings of The 17th International Congress of Phonetic Sciences*, Hong Kong, pp. 148-151.
Xu, C. X. and Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association* **33**: 165-181.
Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics* **25**: 61-83.
Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* **55**: 179-203.
Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* **27**: 55-105.
Xu, Y. (2004). The PENTA model of speech melody: Transmitting multiple

communicative functions in parallel. In *Proceedings of From Sound to Sense: 50+ years of discoverries in speech communication*, Cambridge, MA: C-91-96.

Xu, Y. (2011). Post-focus compression: Cross-linguistic distribution and historical origin. In *Proceedings of The 17th International Congress of Phonetic Sciences*, Hong Kong, pp. 152-155.

Xu, Y., Chen, S.-w. and Wang, B. (in press). Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family? To appear in The Linguistic Review.

Xu, Y. and Sun, X. (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 Wallace, A. (2004). Multiple effects of consonant manner of articulation and intonation type on F0 in English. *Journal of the Acoustical Society of America* 115, Pt. 2: 2397.

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

Xu, Y. and Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33: 159-197.

Zerbian, S., Genzel, S. and Kügler, F. (2010). Experimental work on prosodically-marked information structure in selected African languages (Afroasiatic and Niger-Congo). In *Proceedings of Speech Prosody 2010*, Chicago: 100976:1-4.