Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Journal of Phonetics 33 (2005) 159–197

Journal of Phonetics

www.elsevier.com/locate/phonetics

Phonetic realization of focus in English declarative intonation

Yi Xu[a,b,*], Ching X. Xu[c]

[a]*Department of Phonetics and Linguistics, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, UK*

[b]*Haskins Laboratories, New Haven, Connecticut, USA*

[c]*Department of Communication Sciences and Disorders, Northwestern University, Evanston*

Received 18 July 2003; received in revised form 4 August 2004; accepted 8 November 2004

Abstract

The present study investigates how focus is phonetically realized in declarative sentences in American English. The goal is to test the hypotheses that, (a) focus is manifested *in parallel* rather than *in alternation* with other intonational functions, and (b) every syllable in a sentence is associated with a local pitch target. Eight native speakers of American English recorded short declarative sentences with narrow focus at different locations or without any narrow focus. Detailed f_0 analyses reveal that a narrow focus is realized by expanding the pitch range of the on-focus stressed syllables, suppressing the pitch range of postfocus syllables, and leaving the pitch range of prefocus syllables largely intact. Focus is not found, however, to determine the presence or absence of f_0 peaks. Data analyses also reveal evidence for the presence of a local pitch target in every syllable. These findings are incompatible with conventional theories of English intonation. As an alternative, the Parallel Encoding and Target Approximation (PENTA) model is considered. The model defines and organizes the intonational components in terms of function rather than form. It also assumes target approximation rather than interpolation as the basic articulatory mechanism of f_0 contour generation. It is argued that the approach used in the PENTA model, which takes account of both communicative functions and articulatory implementation, may provide a coherent account of detailed f_0 contours in English.

[*] Corresponding author. Department of Phonetics and Linguistics, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, UK. Tel.: +44 (0) 20 7679 5011.

E-mail address: yi@phon.ucl.ac.uk (Y. Xu).

0095-4470/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2004.11.001

1. Introduction

> Ordinarily, when we speak of stressing something, we refer to giving it special emphasis. But it is misleading to think of the stressed syllable of a word as something that is regularly more emphatic than the other syllables. Rather, that syllable is the one that will get the special emphasis whenever the word is emphasized. (Bolinger, 1986, p. 14)

Focus, which is equivalent to emphasis in the above quote, is a communicative function known to be mainly manifested through f_0 variations (cf. Ladd, 1996 and references therein). The quote by Bolinger expresses a widely received view about how focus is realized in English intonation: *It gives prominence to the syllables that are lexically stressed, primarily by assigning them a pitch accent*. The present study is designed, as its primary goal, to reevaluate this view. The secondary goal, which is necessitated by the first, is to understand the detailed f_0 contours and their alignment with segmental materials as related to focus. We will start with a brief overview of how focus is treated in the two most influential theoretical frameworks of English intonation: the British nuclear tone tradition and the American autosegmental-metrical (AM) framework, both of which share Bolinger's view about emphasis as quoted above.

1.1. The nuclear tone tradition

In the nuclear tone tradition (Crystal, 1969; O'Connor & Arnold, 1961; Cruttenden, 1997; Palmer, 1922), to analyze the intonation of an utterance, for each intonation group a nucleus is first identified as "the stressed syllable of the last prominent word in a sense group" (O'Connor & Arnold, 1961, p. 271). While a variety of nuclear tones have been described, the one most closely related to focus in short declarative sentences is the high-fall (Cruttenden, 1997, p. 51), also known as the High Fall (O'Connor & Arnold, 1961, p. 13). The f_0 contour in the syllable following the nucleus is referred to as the tail (O'Connor & Arnold, 1961), or nuclear tail (Crystal, 1969), or simply as part of the nuclear tone (Cruttenden, 1997). Beside the nucleus and tail, pitch accents before the nucleus are also identified "by an obtrusion of the pitch on one syllable from the pitch on surrounding syllables…" (Cruttenden, 1997, pp. 47–48). These accents are referred to as either the prenuclear accents (Cruttenden, 1997) or the head (Crystal, 1969; O'Connor & Arnold, 1961). A declarative sentence with a narrow focus is therefore described as having a high-fall nucleus, a low flat tail and an unspecified head and/or prehead.

1.2. The AM theory

In the AM theory, unlike in the British tradition where nuclear tones are often described as contours, intonation is described in terms of two level tones—H and L (Beckman & Pierrehumbert, 1986; Ladd, 1996; Pierrehumbert, 1980). It is also assumed that pitch accents, which consist of either a single tone or two successive tones, are "phonologically located on metrically prominent syllables" (Pierrehumbert, 2000, p. 20). Unlike the British tradition which generally recognizes only sense groups (O'Connor & Arnold, 1961), intonation-group (Cruttenden, 1997) or tone-unit (Crystal, 1969), the AM theory assumes that there are two levels of phrasing in English intonation: the intermediate phrase and the full intonational phrase.

An intermediate phrase is marked by a phrase accent at its right (or left) edge (H- or L-), and an intonational phrase is marked by a boundary tone at its right edge (H% or L%).

While the notion of nuclear accent similar to that in the British tradition is also frequently used in the AM theory, it is not recognized as something different from the prenuclear accents other than being the last in an intonation phrase. Instead, it is argued that "the accent inventory is the same in prenuclear and in nuclear position, with the more complex configurations found in nuclear position being attributable to extra tones originating at the phrasal level." (Pierrehumbert, 2000, p. 26). Thus the AM theory treats a declarative sentence with a narrow focus as having a $L + H^*$ (or H^*) pitch accent followed by a L- phrase accent and a L% boundary tone.

1.3. Form versus function

Despite their differences, the nuclear tone tradition and the AM theory share one important characteristic. That is, in both systems the components of intonation and their organization are defined *primarily* in form, while the meanings of these components are assigned only after their forms are established. The form of an intonational component involves two aspects. The first is its shape and relative height. In the AM approach, for example, a H is identified as corresponding to an f_0 peak while a L an f_0 valley. The second formal aspect of a tonal component is its relative prominence. This is mainly determined by auditory impression in both approaches (Cruttenden, 1997; Wightman, 2002).

The form-oriented approaches used in the British tradition and AM theory are rather different from those taken in the study of lexical tones. Lexical tones are defined not first by form, but by function: they are recognized as serving to distinguish words or morphemes that are segmentally identical. It is *after* this property has been established when instrumental investigation (Bai, 1934; Liu, 1924) as well as auditory assessment (e.g., Chao, 1930, 1956) are carried out to find the physical correlates of the distinctive property.

There is no doubt that the lack of lexical tone in a language like English makes a function-oriented approach to intonation rather difficult. But the difficulty does not mean that the approach is not worth pursuing. After all, in any language, including English, it is unlikely that communicative functions have evolved to serve preexisting phonological forms. Rather, it is more likely that phonological forms have evolved to serve various communicative functions. It is therefore possible that the understanding of English intonation can be improved by considering its components as directly related to communicative functions. To overcome the apparent difficulty just mentioned, we can start with a relatively salient function—*focus*, whose acoustic manifestation has been found to be almost as robust as lexical tone in Mandarin (Xu, 1999), and whose close link to certain salient acoustic patterns in English has been widely recognized, as will be discussed next.

1.4. Focus as a communicative function

From a functional perspective, focus refers to an emphasis on some part of a sentence as motivated by a particular discourse situation. For example, a narrow focus is put on "red" when "Your eyes are red" is said in response to "What's wrong with my eyes?" In contrast, the focus will be on "eyes" when the question is "Is my nose red?" In this sense, *focus is a discourse function* *serving to highlight a particular piece of information against information already shared by the conversation participants* (see Bolinger, 1972; van Heuven, 1994; Ladd, 1996 and Gussenhoven, in press for more detailed discussion). Note that in this functional "definition" nothing is said about what the phonetic or phonological form of the highlighting is. In contrast, in the nuclear tone tradition, "EYES" when focused would be described as having a high-fall nuclear pitch accent. Similarly, in the AM theory, "EYES" would be assigned a pitch accent, likely L + H*, followed by a L- phrasal tone. Thus in both frameworks, focus manifests itself via a process of placing nuclear pitch accents, also known as "Focus-to-Accent" assignment (Gussenhoven, 1985; Ladd, 1996). Critically, once a word or syllable receives the nuclear accent, that accent is the only intonational component it carries. Furthermore, in both theories, by definition, there are no more prominence-related pitch events after the nuclear accent.

This appears to be very different from the realization of focus in tone languages. As found by Gårding (1987), Shih (1988), Jin (1996), Xu (1999) and Chen (2003), in Mandarin a narrow focus neither replaces nor deletes the lexical tones. Instead, as two separate functions, tone and focus are concurrently realized by varying different aspects of f_0 contours. Tones are implemented as local pitch targets, while focus as regional pitch range variations. This can be seen in Fig. 1a, where the pitch range directly under focus is expanded; the pitch range after focus is suppressed (lowered and compressed); and the pitch range before focus is virtually identical to that of a sentence with no narrow focus. Similar concurrent implementation of focus and lexical tone has been reported for Shanghai Chinese (Selkirk & Shen, 1988) and Cantonese (Man, 2002).

Focus-related f_0 variations in Mandarin, however, can be also seen as similar to focus realization in English as described by both the British tradition and AM theory. For example, in Fig. 1b and c, where focus is on the first disyllabic word, if we single out any individual curve without comparing it directly to others, only a single prominent f_0 peak is clearly visible, after which hardly anything can be discerned as fully distinct. But if we do compare the three curves directly, as done in these plots, the differences due to the lexical tones are still visible, albeit much reduced. Focus realization in tone languages therefore tells us that, by only making pitch range adjustments, a focus neither replaces the tone directly under it, nor eliminates the tones after it. It also tells us that tone and focus can be both transmitted through f_0, but they each modify f_0 in different ways.

Given that even in a tone language like Mandarin, as seen in Fig. 1b and c, focus can effectively mask the (visual) trace of lexical tones when an f_0 track is inspected in isolation, it is possible that similar processes happen in English as well. Thus there is a need to reexamine if focus is realized *in alternation* or *in parallel* with other intonational functions. The likely other functions in English include lexical stress and rhythm. The former serves to distinguish words and the latter probably helps to demarcate syllable/word strings into larger chunks.

1.5. f_0 contours and their alignment with syllables

The f_0 plots in Fig. 1 further show that focus affects not only the f_0 height of syllables in Mandarin, but also their f_0 contour shapes. But both effects can be understood if tone and focus are viewed as separate functions (Xu, 1999). For English, neither the British tradition nor the AM theory specifies the exact alignment of f_0 contours with the segmental material. As a result, several questions remain open.

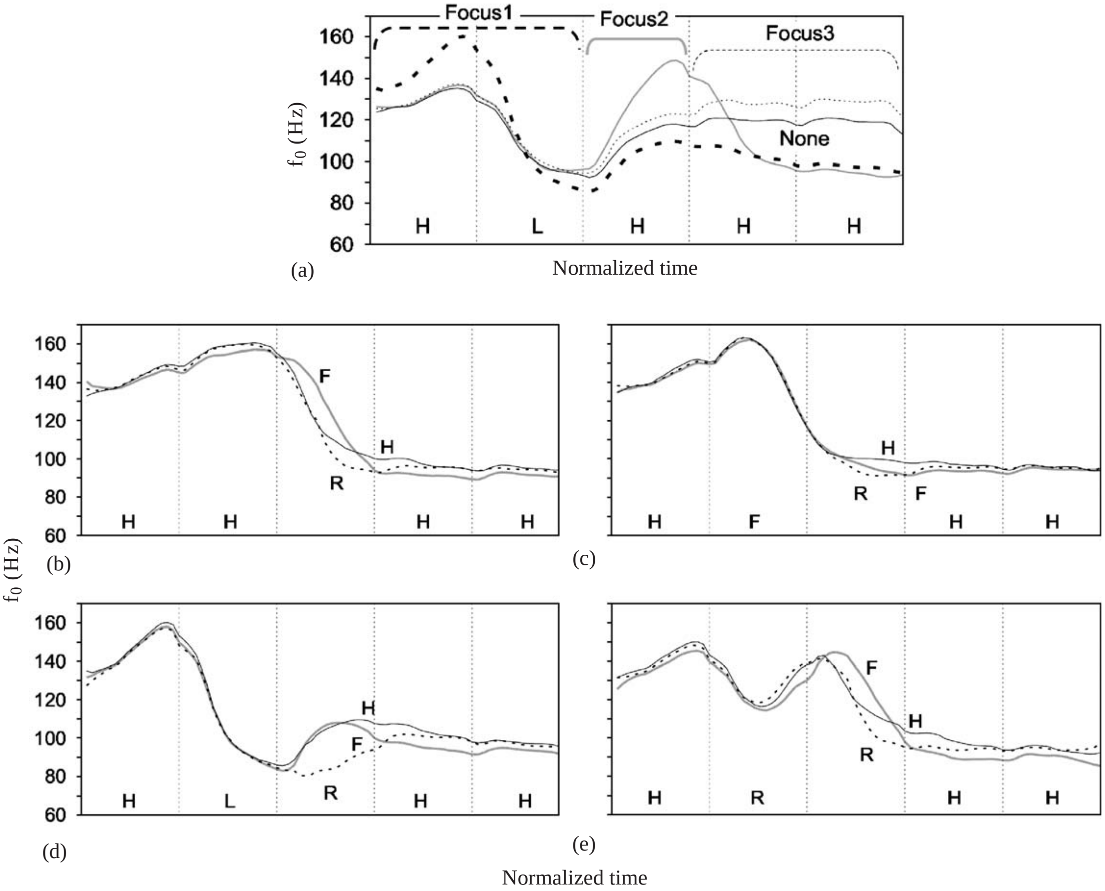


Fig. 1. (a) Interaction of tone and focus in Mandarin H H H H H (left) and H L H L H (right) sequences. The location and temporal scope of focus are indicated by the labels near the curves. (b) Suppression of postfocus tones when syllable 2 carries four different tones. In all cases, the focus is on the first two syllables. All plots adapted from Xu (1999).

First, why should f_0 peaks and valleys be where they are? The AM theory specifies that a starred tone is aligned with a lexically stressed syllable. But it provides no mechanistic explanation for the exact alignment of f_0 peaks and valleys with that syllable. While there have been studies on the actual alignment of f_0 peaks and valleys and the factors affecting it (Arvaniti, Ladd, & Mennen, 1998; Silverman & Pierrehumbert, 1990; Ladd, Faulkner, Faulkner, and Schepman, 1999; Ladd & Schepman, 2003, etc.), no mechanistic explanations are offered. A recently emerging account is that certain f_0 peaks and valleys are "anchored" to the segmental locations exactly as they are observed (Arvaniti et al., 1998; Atterer & Ladd, 2004; Ladd et al., 1999; Ladd & Schepman, 2003). In other words, observed alignment *is* phonological alignment.

Second, why should f_0 contours have the exact shapes that they have? The AM theory again provides rules regarding the shapes of the pitch accents. In fact it is the shape considerations that have motivated the proposal of double-toned accents such as L + H* as opposed to H* and L*

(Pierrehumbert, 1980). But again there are no mechanistic explanations, and the shape definitions are understood as directly reflecting the underlying forms of the intonational components. In other words, observed contour *is* phonological contour.

Third, what gives rise to the f_0 shapes of the nonaccented syllables and words? Here the AM theory does provide a mechanistic explanation: their f_0 values come from interpolation between surrounding pitch accents; and the interpolation is either linear, or nonlinear with a "sagging" function (Pierrehumbert, 1980, 1981). This, in our view, is a critical starting point for understanding the mechanisms of intonation. As has been argued forcibly in many occasions (e.g., Beckman, 1995; Ladd, 1996; Pierrehumbert, 1980), the AM theory is strictly linear. That is, "for any given target tone, the implementation was held to depend only on the identity and prosodic position of the tone itself, and on the identity and phonetic realization of the preceding tone" (Pierrehumbert, 2000, p. 29).

Note that this assumption of strict phonetic linearity is in direct conflict with the interpolation mechanism assumed in the theory, because interpolation necessarily *entails* anticipation, as illustrated in Fig. 2a. According to Pierrehumbert (1981), pitch accents (illustrated by the two peaks labeled H*) are first *assigned precise numerical values* by phonetic implementation rules. After that, the f_0 between the two peaks are derived with either a linear interpolation function such as $f_0 = at + b$, or a nonlinear, parabolic (i.e., sagging) function: $f_0 = at^2 + bt + c$. Note that this derivation process is precisely doing "phonetic lookahead," as explained in Pierrehumbert (1981, p. 992): "… ours is the only one in which the upcoming peak affects the f_0 contour from the moment it leaves the last peak." This theory internal conflict thus casts doubt on interpolation as a basic mechanism for generating detailed surface f_0 contours.

1.6. An articulatory perspective

As found in recent research, articulatory movements in speech are often as fast as speakers can possibly make (Janse, 2003; Xu & Sun, 2002). In the case of pitch production, Xu and Sun (2002) compared the maximum speed of pitch change obtained in a forced imitation task to the maximum speed of pitch change reported for several languages, including Mandarin, English, and Dutch. They found the two largely comparable in all these languages. The finding suggests that the constraint of the maximum speed of pitch change plays a significant role in shaping the f_0 contours in speech. For example, according to Table V of Xu and Sun (2002, p. 1407), it takes over 100 ms for an average speaker to raise or lower pitch by just 2 semitones. This means that transitions from one intended pitch level to another make up most of the surface f_0 contours. On the other hand, precisely because the nature of a transition is to approach an intended value, a surface f_0 configuration necessarily becomes more and more like the targeted form as time passes. By the same mechanism, no matter how different the starting points are, transitions to an underlying form should gradually converge to its ideal configuration. This convergence property can then be employed to reveal the ideal configuration of an underlying form, as has been done for Mandarin (Xu, 1997, 1999). Xu and Wang (2001) refer to the converged f_0 configurations as the underlying *pitch targets* of the corresponding tones. Note that this notion of "target" is different from that of the "tonal target" in the AM theory, because the

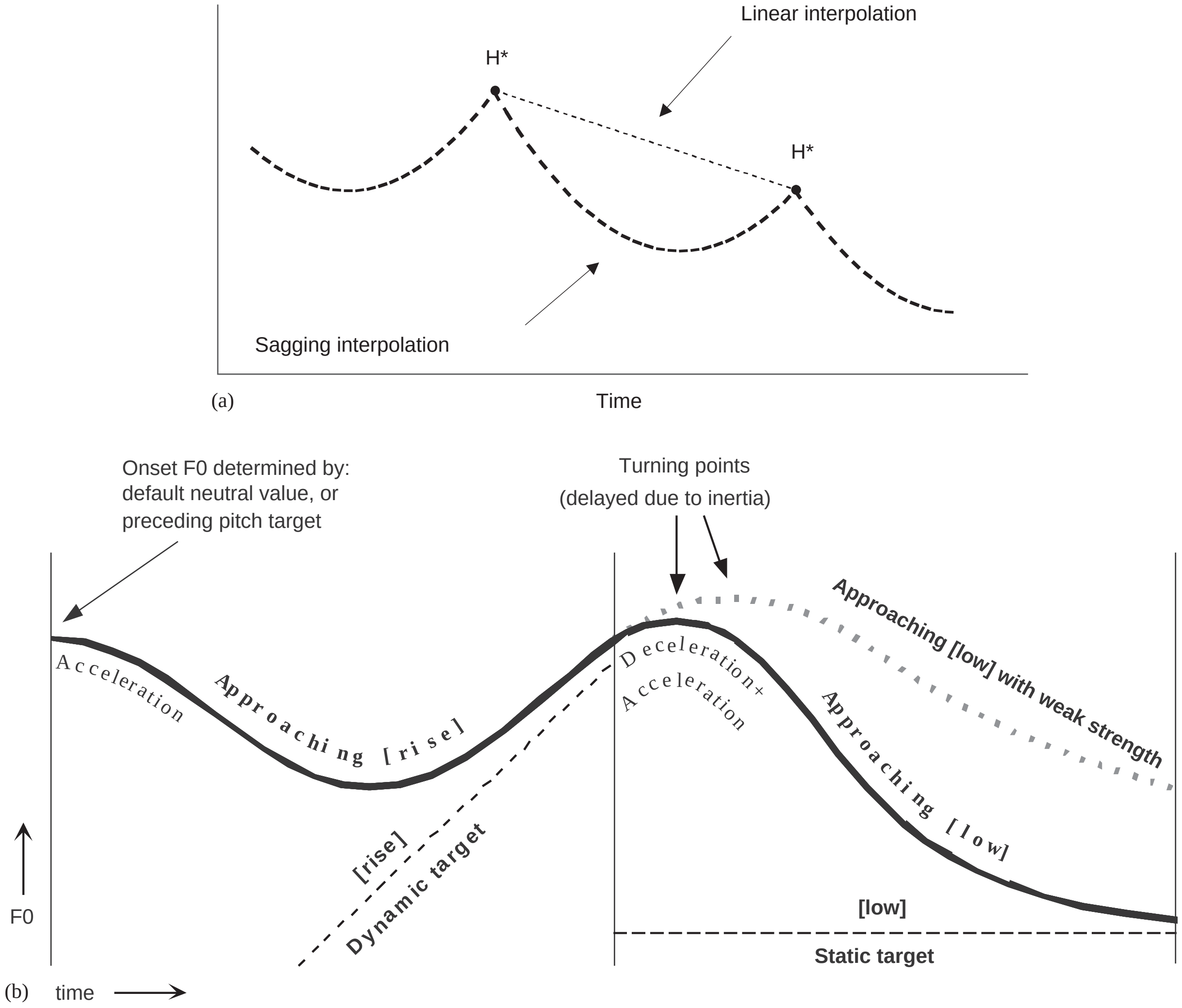


Fig. 2. (a) Illustration of linear and "sagging" interpolation proposed by Pierrehumbert (1980, 1981). (b) Illustration of the Target Approximation model. The vertical lines indicate syllable boundaries. The straight dashed lines represent *local pitch targets*. The solid curve depicts the f_0 contour resulting from *asymptotic approximation* of the pitch targets. (Adapted from Xu & Wang, 2001.) The dotted curve in syllable 2 simulates the effect of **weak** articulatory strength.

former refers to articulatory goals that are *covert*, while the latter to the *actual* peaks or valleys in surface f_0 contours.

Two additional observations about the f_0 convergence in tone production are also relevant to the present study. The first is that for tones such as R (Rising tone) and F (Falling tone) in Mandarin, the converged configurations are slopes rather than horizontal lines as in the case of H (High tone). This suggests that the ideal f_0 configurations in these cases are dynamic rather than static. The second is that for each tone the transitions always start roughly from the onset of the syllable and end around the offset of the syllable, regardless of whether the converging configuration is a slope or a horizontal line. Thus for Mandarin at least, the f_0 transition toward the underlying pitch target of a tone appears to coincide with the syllable.

Data on tonal realization in Mandarin therefore suggest *syllable-synchronized sequential target approximation* as the mechanism of tonal implementation (Xu & Wang, 2001). An illustration of this mechanism is shown in Fig. 2b, in which the straight dashed lines represent *local pitch targets*. During target approximation, at each moment in time, the present articulatory state is compared only to the desired state for the *current* target, and the difference between the two determines the direction and speed of further f_0 movement. There is thus neither lookahead at the next target nor lookback at the previous target. Nevertheless, a strong influence from the previous target naturally occurs because its implementation gives rise to the initial articulatory state for the implementation of the current target. This "carryover" influence, also naturally, diminishes over time as the current target is being approached. The approximation of the next target starts as soon as the current syllable is over, but not any time sooner. Based on these mechanistic suppositions, the validity of interpolation and target approximation can be compared by examining the influences of adjacent syllables on each other's f_0 values. Interpolation would predict both carryover and anticipatory influences. Target implementation would predict, in contrast, predominant but fading carryover influence from the preceding syllable, with little or no influence from the upcoming syllable.

1.7. Research questions

The foregoing literature review has demonstrated the need to reexamine English intonation from two rather different yet intimately related perspectives. First, it is necessary to consider intonational components as communicative functions rather than as just visually and/or auditorily observed forms. From such a perspective, we need to explore whether focus, as a communicative function, is realized *in parallel* or *in alternation* with other intonational functions. Second, it is necessary to consider surface f_0 events as products of articulatory execution rather than as direct replicas of the underlying components proper. From such a perspective, we need to explore whether the basic mechanism of f_0 generation is *interpolation* across observed f_0 peaks and valleys or sequential *approximation* of underlying pitch targets whose forms do not directly resemble surface f_0 contours.

To address these issues, we will examine detailed f_0 contours in short declarative sentences in English said with narrow focus at various locations or without narrow focus. The goal is to find answers to the following two main questions and their corollaries:

(1) Is focus realized in parallel or in alternation with other intonational components in English? More specifically,
    (a) Are there local f_0 movements that are largely independent of focus?
    (b) Are there f_0 patterns that are unique to focus, i.e., largely independent of other factors?
(2) Are the shape and alignment of f_0 contours in English better accounted for in terms of interpolation between accents or sequential approximation of successive underlying pitch targets in each and every syllable? More specifically,
    (a) Are the alignment of f_0 peaks and valleys the result of direct anchoring or consequence of implementing underlying pitch targets?

(b) Do syllables between pitch accents have their own pitch targets, or is f_0 only interpolated through these syllables?

2. Method

The design of the study is to find answers to the aforementioned research questions by performing detailed acoustic analyses of f_0 contours in English. The f_0 analyses to be performed are very detailed, and so it is impractical to include many experiment conditions. We thus restricted the project to only short English declarative sentences spoken with or without narrow focus on different words and at different speaking rates. We also limited the number of combinations of these factors, for otherwise the amount of data would be too massive.

2.1. Stimuli

The stimuli are short declarative sentences. To make extensive f_0 alignment analyses possible, we need to use words that have sonorant (preferably nasal) onsets and with no coda consonants if possible. The target sentences used are in the form of "*Lee* may *know* my *niece*." The italicized words are referred to as the "key words." They vary in word length, stress pattern, phonological length of stressed syllable and focus. Word length varies from monosyllabic to trisyllabic. Lexical stress varies between word-final (including monosyllabic words) and non-final. Phonological length of stressed syllable is either long or short. Focus varies from sentence-initial (word 1), sentence-medial (word 2), sentence-final (word 3) to neutral focus (i.e., no narrow focus). Due to the kind of detailed acoustic analysis to be performed, only a limited number of combinations of these factors will be examined.

The following are the compositions of the stimulus sentences. The words "may" and "my" remain unchanged in all sentences, and they are referred to as "non-key words." They are also usually unaccented in the conventional sense both in the British tradition and in the AM theory (being either an auxiliary verb or a personal pronoun, unless in special contexts, which are not included in the present design). Three sentence groups are composed for examining f_0 contours at three locations in the sentence: beginning, middle, and end. In each sentence group, the alternative words in the same location rotate to form different sentences. Sentences in each group were produced in two focus conditions: neutral focus, and focus on the underscored word.[1]

1. Lee/Nina/Lamar/Emily/Ramona may know my niece — 5(words) × 2(foci) × 7(repetitions) = 70
2. Lee may lure/mimic/minimize my niece — 3(words) × 2(foci) × 7(repetitions) = 42
3. Lee may know my niece/nanny/mummy — 3(words) × 2(foci) × 7(repetitions)−7 = 35[1]

Focus is controlled by having subjects say the target sentences as answers to prompt questions that ask about specific pieces of information available in the target sentences. This method has been used successfully in previous studies (Cooper, Eady, & Mueller, 1985; Xu, 1999). The prompt

[1] The "−7" is because unfocused "Lee may know my niece" is used to contrast both with focused "Lee" and with focused "niece," as shown below.

questions are shown below together with illustration of focus locations in exemplar target sentences.

Prompt:	Target:
Who may know your niece?	*Lee may know my niece.*
What may Lee do to your niece?	*Lee may lure my niece.*
Who may Lee know?	*Lee may know my niece.*
What did you say?	*Lee may know my niece.*

To ensure the reliability of the regression analyses on f_0 alignment a wide range of duration variations need to be elicited. This was done by having subjects say the same sentence at two speaking rates: normal and fast. (A pilot test showed that some speakers had difficulty maintaining focus consistently at slow speaking rate. So, only two speaking rates were used.)

2.2. *Subjects*

Eight native speakers of American English, aged 20–35, participated as subjects. Four of them were female, and the others male. They were recruited from the Northwestern University campus and were paid for their participation. None of them reported having any speech disorders. They all spoke general American English without noticeable accents.

2.3. *Recording procedure*

Recording was conducted in a sound-treated booth at the Speech Acoustics Laboratory in the Department of Communication Sciences and Disorders at Northwestern University. The subject was seated comfortably in front of a computer monitor. The microphone was placed by the side of the monitor, approximately 1 foot away from the subject's lips. In each trial, the subject pressed the "Next" button displayed on the screen and the target sentence was displayed on the screen. At the same time, a prompt question was played through a loudspeaker. The subject then read aloud the displayed sentence as a response to the prompt question. The prompt questions were recorded at two speaking rates, normal and fast. Subjects were instructed to say the target sentence at a similar speaking rate as that of the prompt question. They were also instructed not to pause in the middle of a sentence. In case a mistake was made as judged by the experimenter, the subject was asked to repeat the sentence. The sentences were presented in random order, and a different order was used for each subject. Before the start of the real trials, the subject went through a number of practice trials until he/she was familiar with the procedure.

2.4. *f_0 extraction*

The acoustic analysis procedure was similar to those used in Xu (1997, 1998, 1999, 2001a). First the digitized signals were converted to a format readable by programs in the ESPS/waves+ signal processing software package (Entropic Inc.). Then individual target sentences were extracted and saved as separate ESPS signal files. The program *epochs* in the ESPS package was then run to mark every vocal cycle in the target words. After that, the marked signals were labeled manually in the ESPS *xwaves* program for the onset and offset of each segment (both consonants and vowels) of the target words using the *xlabel* program. Manual editing was performed to correct spurious vocal pulse labeling by the *epochs* program (such as double-marking or vocal-cycle skipping).

The vocal pulse markings and segment labels for each utterance were saved by the *xlabel* program into a text file. Those text files were then processed by a set of custom-written computer programs. These programs first converted the duration of vocal cycles into f_0 values, and then smoothed the resulting f_0 curve using a *trimming algorithm* to eliminate abrupt bumps and sharp edges (cf. Xu, 1999 for details).

2.4.1. Exclusion of outliers

After extracting individual f_0 curves, we checked all of them for outliers. The purpose was to exclude sentences that were said with apparently wrong focus. Exemplar f_0 contours of the seven repetitions of each sentence with the same focus and speaking rate are shown in Fig. 3, which displays the f_0 contours of the sentence "Nina may know my niece" with no narrow focus, produced by all subjects at "normal" rate. These curves are displayed using normalized time, i.e., with the same number of points taken from each syllable at equal proportional intervals, e.g., 0, 1/20, 2/20, 3/20, …, 20/20. As can be seen, displayed in this way, the f_0 curves by each subject, except subject 2 (whose case will be discussed later), are highly consistent across the seven repetitions. When an inconsistency was noticed, the following criteria were used to determine if an outlier was involved and if it should be excluded.

A repetition is excluded if and only if

(a) it is obviously different from the rest of the repetitions, and
(b) it has the wrong focus as judged auditorily by the authors

A case in point is the curve in row 2, column 1 that drops to the bottom around the middle of the sentence and remained low throughout the rest of the sentence. That sentence clearly sounded as if there was a sentence-initial focus.

A repetition is not excluded if:

(a) it differs from other repetitions only in pitch range but not in perceived focus, or
(b) it seems to differ from other repetitions in melodic pattern (in the case of subject 2)

Altogether, a total of four repetitions from subject 2 were excluded, and 1 from subject 4 was excluded.

2.4.2. Grand mean f_0 curves for visual inspection

After excluding the outliers, for each subject, the time-normalized f_0 curves of individual repetitions of each sentence in each focus condition at each speaking rate were averaged to obtain a mean f_0 curve. Then the mean duration of each syllable across the repetitions was computed. This mean duration was used in displaying the f_0 contours of each syllable in the sentences in the same focus condition at the same rate. In this way we could compare the tonal contours of different sentences without losing sight of the actual duration of each syllable. Fig. 4 displays mean f_0 curves of all sentences produced at normal rate by all

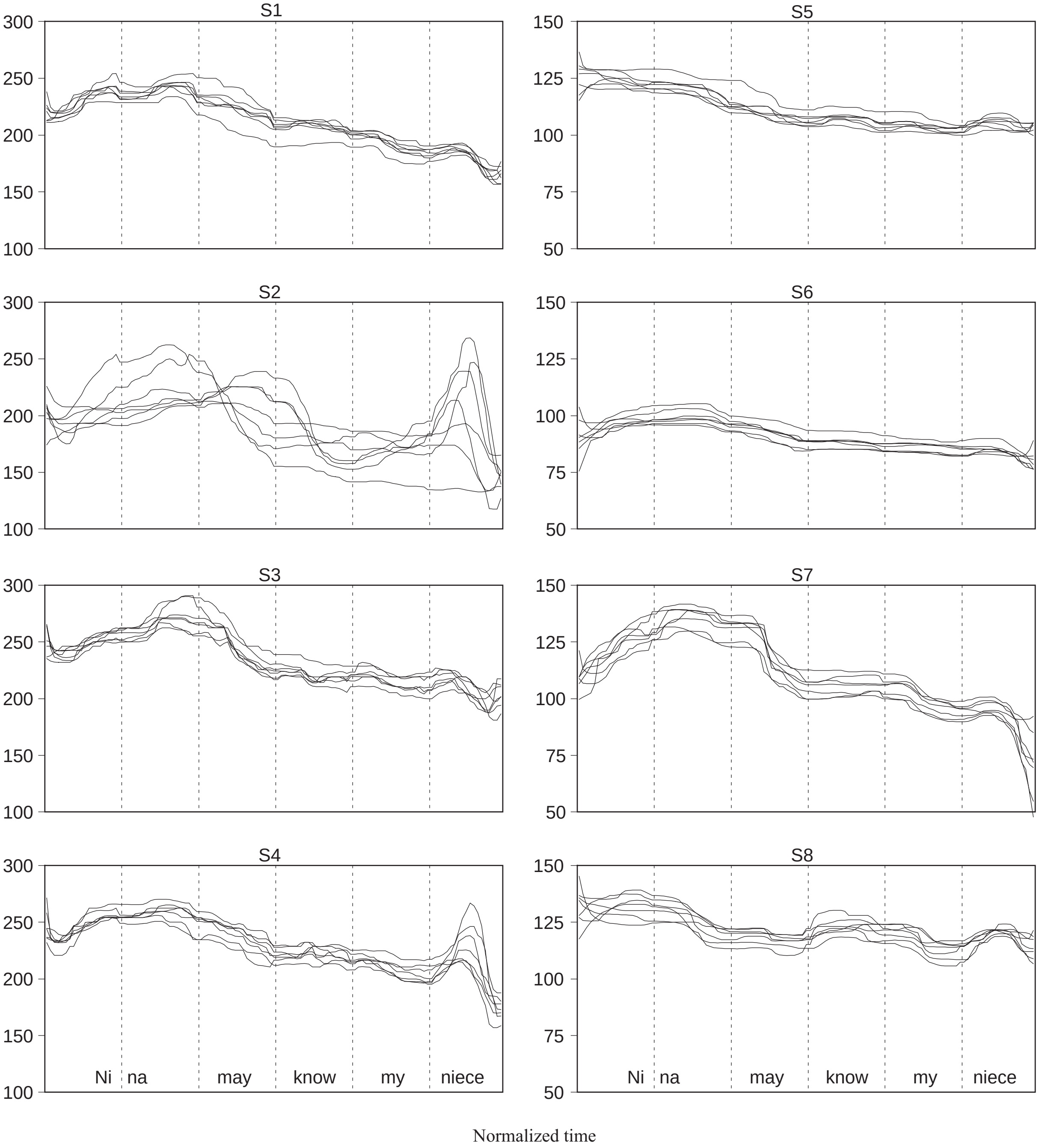


Fig. 3. Time-normalized f_0 curves of seven repetitions of “Nina may know my niece” said by eight subjects at “normal” rate with no narrow focus. Subjects 1–4 are female; and subjects 5–8 are male.

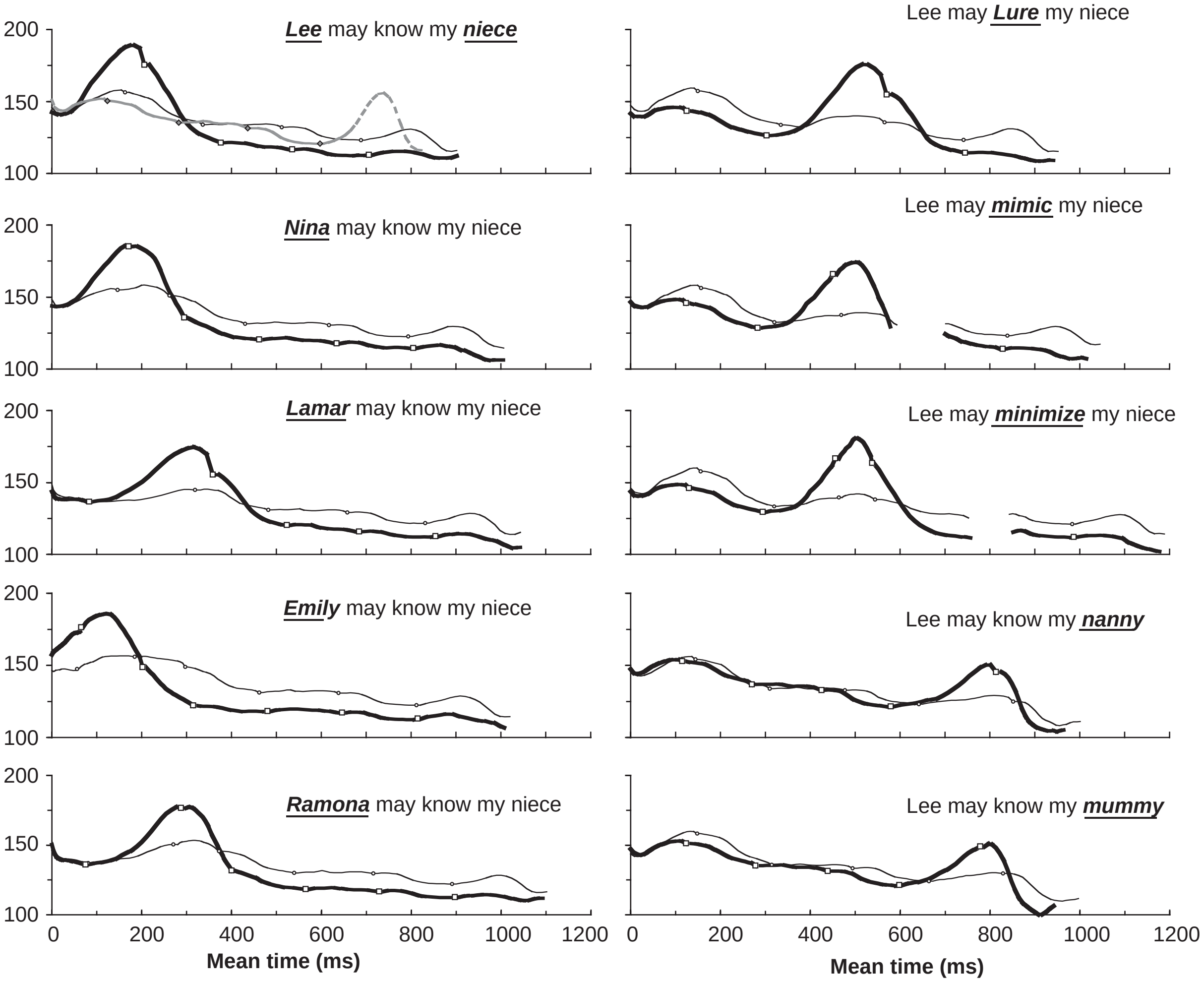


Fig. 4. Mean f_0 contours of all sentences produced at normal rate by 7 subjects. In each graph, the ordinate is the mean f_0 in Hz averaged over 49 repetitions by 7 subjects, and the abscissa is time in ms. The duration of each syllable in a f_0 curve is the grand average of 49 repetitions by 7 subjects. The thicker curves have narrow focus on one of the words as indicated by the underscore in the sentence printed in each graph. The open squares and circles indicate syllable boundaries, located at the first vocal pulse of the initial consonants. In the sentences containing the words "mimic" and minimize," the gaps in f_0 curves correspond to the closure or frication of the final consonants.

subjects except subject 2. The f_0 curves of subject 2 were not included in the mean f_0 curves because of their apparent inconsistencies with those of other subjects'. The open squares, circles and diamonds on the f_0 curves indicate syllable boundaries. For syllables with initial sonorants, the boundaries are set at the point where the spectral pattern makes an abrupt shift into a typical nasal or lateral pattern. (cf. Xu, 1999 for more detailed description of the labeling procedure). For syllables with stops and fricatives, the boundaries are set at the onset of closure or frication.

2.5. Measurements

Listed in the following are measurements taken from f_0 curves produced by all subjects using a set of custom-written C programs. They were taken from individual trimmed f_0 curves rather than from the mean f_0 curves averaged across subjects or across repetitions. Some of the measurements are intermediate ones used only in the calculation of other measurements. For the purpose of this study, a syllable is defined as consisting of an onset consonant, which in the case of the first syllable in "Emily" is a glottal stop, a vowel or diphthong and an optional coda (in the case of final syllables in "mimic", "minimize" and "niece"). Acoustically, the beginning of consonant closure is treated as the syllable onset, and the end of vowel (when there is no coda) or the end of release of the coda is treated as the syllable offset.[2]

- Maxf_0 (st)—highest f_0 in the stressed syllable of the key words, measured in semitone with the lowest f_0 of each subject as the reference, for assessing f_0 peak height as well as pitch range.
- Minf_0 (st)—lowest f_0 in the stressed syllable of the key words (or in all words for some analyses), measured in semitone with the lowest f_0 of each subject as the reference, for assessing lowest f_0 as well as pitch range.
- Rise size (st)—difference in semitone between maximum f_0 and minimum f_0 in the stressed syllable of a key word, for assessing pitch range.
- Rise time (ms)—time interval between f_0 minimum and maximum in the stressed syllable of a key word, an intermediate measurement for calculating rise speed.
- Rise speed (st/s) $= 1000 * \text{Rise size}/\text{Rise time}$, for assessing the effect of focus.
- Stress-dur (ms)—duration of the stressed syllable of a key word, for assessing lengthening by focus.
- C1-to-Minf_0—time interval between onset of the stressed syllable of a key word, where C1 = closure onset of sonorant consonant (/l/, /m/ and /n/ in stressed syllable and closest f_0 minimum. This is for assessing f_0 valley location relative to onset of stressed syllable.
- Maxf_0-to-C2—time interval between onset of the first poststress syllable and closest f_0 maximum, where C2 = closure onset of nasal consonant following the stressed syllable of key word (not applicable in word 3 with final stress). This is for assessing f_0 peak delay beyond a stressed syllable.
- C1-to-maxf_0—time interval between onset of the stressed syllable of a key word and closest f_0 maximum; an intermediate measurement for calculating peak location.
- Peak location = 100 × C1-to-maxf_0/Stress-dur, for assessing f_0 peak location relative to onset of stressed syllable of a key word.
- Valley location = 100 × C1-to-minf_0/Stress-dur, for assessing f_0 valley location relative to onset of stressed syllable of a key word.

[2] Recent evidence has suggested, however, this is probably not the most accurate definition of the syllable from an articulatory perspective, cf. Xu and Liu (2002) and Liu and Xu (2003), although the current definition is sufficient for the purpose of this study.

3. Analyses and results

The following analyses were performed to help us find answers to the two main questions and their corollaries raised in the Introduction. The analyses consist of three general procedures. First, through visual inspection of the f_0 contours we observed the gross pattern of f_0 variation in height and alignment as related to focus and word stress. Second, a set of repeated measures ANOVAs were performed to analyze the gross pattern of focus realization (as well as the alignment of f_0 valley). Third, a set of linear regressions were performed on the detailed alignments of f_0 peaks and valleys, with the goal of identifying the sources of the alignment patterns. In the ANOVAs, speaking rate was not used as a factor, because its effect is largely predictable and not essential for the purpose of those analyses. The measurements used in the ANOVAs are therefore averages across the two speaking rates, normal and fast. The measurements are also averaged across repetitions, and for each repeated measures ANOVA subject is always the random factor. For the regression analyses, however, gradient variations introduced by speaking rate manipulation are essential for the reliability of analysis.

3.1. Focus effects

When the mean f_0 contours of the same sentence uttered with and without a narrow focus are superimposed on each other, the main effects of focus become quite evident. A number of observations are listed in the following, which are all visible in Fig. 4.

1. The f_0 peak of a word is consistently higher under a narrow focus than in the neutral-focus sentence. At the same time, the general locations of the f_0 peaks are largely the same with or without narrow focus. This is seen in every graph in Fig. 4.
2. The f_0 peaks of *all* postfocus words are lower than those of the same words in the neutral-focus sentence. Nonetheless, there are also some visible f_0 peaks corresponding to the key words in the postfocus region. This is seen in all graphs where there is a nonfinal focus.
3. The f_0 peaks of prefocus words are lower than those of the same words in the neutral-focus sentence for some subjects but not for others. This is reflected by the mean f_0 curves in the first three graphs in the right column. The same tendency can also be seen in the first graph of the left column and the last graph of the right column.
4. The scope of postfocus suppression seems to include not only all postfocus words but also the final unstressed syllable(s) in the focused word: "Nina", "Emily", "Ramona", "mimic", "minimize", "nanny" and "mummy".

To verify the visual observations, a set of repeated measures ANOVAs were performed. Table 1 displays *max*f_0, *min*f_0, *rise size, rise speed, and stress-dur* broken down according to focus (on/none), lexical stress (word-final/non-final), and word position (word1/word3/word5). Also displayed in the table are probability values resulting from 3-factor repeated measures ANOVAs performed on the five measurements. To adjust for potential significance inflation due to multiple comparisons, besides the commonly-used probability levels of 0.05, another level of significance was computed using the Bonferroni adjustment: $p = 0.05/5 = 0.01$. (The effect of gender was found to be non-significant for any of the dependent variables in a set of 4-factor mixed-measure

Table 1
Mean values of various measurements under the effects of focus, rate, lexical stress and position, together with probability values from 3-factor repeated measures ANOVAs

	Focus		Lexical stress		Position		
	Yes	No	Word-final	Nonfinal	Word 1	Word 3	Word 5
Maxf$_0$ (st)	11.0	8.2	9.6	9.7	10.9	9.9	8.1
	$F(1,6) = 14.73$,	$\boldsymbol{p<0.01}$	$F(1,6) = 2.80$,	*NS*	$F(2,12) = 32.14$,	$\boldsymbol{p<0.0001}$	
Minf$_0$ (st)	6.6	6.8	6.5	6.9	7.7	6.8	5.6
	$F(1,6) = 1.38$,	*NS*	$F(1,6) = 75.55$,	$\boldsymbol{p<0.001}$	$F(2,12) = 25.47$,	$\boldsymbol{p<0.0001}$	
Rise size (st)	4.4	1.4	3.0	2.8	3.2	3.0	2.4
	$F(1,6) = 14.27$,	$\boldsymbol{p<0.01}$	$F(1,6) = 5.38$,	*NS*	$F(2,12) = 2.52$,	*NS*	
Rise speed (st/s)	23.4	9.5	17.8	15.1	18.2	16.6	14.5
	$F(1,6) = 20.18$,	$\boldsymbol{p<0.01}$	$F(1,6) = 113.44$,	$\boldsymbol{p<0.0001}$	$F(2,12) = 4.46$,	$P<0.05$	
Stress duration (ms)	222.6	195.4	242.3	175.7	188.5	208.0	230.5
	$F(1,6) = 79.53$,	$\boldsymbol{p<0.001}$	$F(1,6) = 198.15$,	$\boldsymbol{p<0.0001}$	$F(2,12) = 224.15$,	$\boldsymbol{p<0.0001}$	

p values smaller than 0.01 (after Bonferroni adjustment for multiple comparisons) are printed in boldface.

ANOVAs. We therefore excluded it in the ANOVAs reported in Table 1.) As can be seen in Table 1, the effect of focus is highly significant for all dependent variables except minf$_0$. Under focus, maximum f$_0$ becomes higher, the size of f$_0$ rise becomes larger, the speed of f$_0$ rise becomes faster, and the duration of the stressed syllable becomes longer. (This is consistent with many previous findings, e.g., Mandarin: Chen, in press; Xu, 1999; English: Cooper et al., 1985; Swedish: Heldner & Strangert, 2001). It is worth pointing out that although the speed of f$_0$ rise under focus increases drastically, it is still well below the maximum speed of pitch rise reported by Xu and Sun (2002) for the corresponding rise size (23.4 st/s at 4.4 st versus $10.8 + 5.6 \times 4.4 = 35.4$ st/s per Table VI in Xu & Sun, 2002). But the speed is similar to what was reported by Ladd et al. (1999) and Ladd, Mennen, and Schepman (2000).

Table 1 shows that the main effect of lexical stress is significant on minf$_0$, rise speed and stress duration. It is not significant on maxf$_0$, but there is a significant interaction between lexical stress and word position (in sentence), which will be explained later. When stress is word-final, the duration of the stressed syllable is increased by 66.6 ms, but the speed of f$_0$ rise is also increased. The increase in rise speed may seem to be related to the increase in rise size, because, according to Xu and Sun (2002), rise speed is directly related to rise size. However, the range of rise size increase in Table 1 is only 0.2 st, which, according to Table VI of Xu and Sun (2002), can generate a speed difference of only 0.72 st/s, much smaller than the 2.7 st/s shown in Table 1. This rise speed increase thus appears deliberate. However, there is a significant interaction between lexical stress and word position [$F(2,12) = 12.94, p<0.01$]. The largest difference between word-final and non-final stress is in word 5 (i.e., in sentence-final words) (5.8 st/s), whereas in word 1 (sentence-initial words) and word 3 (sentence-medial words) the differences are 1.5 and 0.9 st/s, respectively. There is also a significant three-way interaction between focus, lexical stress and word position [$F(2,12) = 9.57, p<0.01$]. The largest difference between word-final and non-final stress is in word 5 under focus: 9.2 st/s, whereas in word 1 and 3 either under focus or not under focus, the largest difference is 2.3 st/s. Thus the sentence final position under focus seems somewhat special. This will become clearer in our further analysis.

The effect of word position is significant for all dependent variables except rise size. As the position of the key word becomes later in a sentence, maximum and minimum f_0 become lower, rise speed becomes slower, and stress duration becomes longer. There are also significant interactions between lexical stress and word position on maxf$_0$ [$F(2,12) = 16.16, p<0.001$], minf$_0$ [$F(2,12) = 13.09$, $p<0.01$], rise speed [$F(2,12) = 12.94$, $p<0.01$], and stress-duration [$F(2,12) = 12.36$, $p<0.01$]. These interactions are not directly relevant to the purpose of the present study, however. So we will not discuss them in detail.

To examine the effect of focus on the postfocus words, a set of 2-factor repeated measures ANOVAs were performed on the f_0 of the postfocus syllables and the results are displayed in Table 2. To adjust for potential significance inflation due to multiple comparisons, besides the commonly-used probability levels of 0.05, another level of significance was computed using the Bonferroni adjustment: $p = 0.05/4 = 0.0125$.

The upper half of Table 2 shows the mean values of maximum f_0 in each word after word 1 and word 3 when they are under focus and when there is no narrow focus in the sentence. As can be seen, the maximum f_0 of words following the focused word is significantly lower than that of the same words in the neutral focus condition whether focus is on word 1 or word 3 (only at $p = 0.05$ level for word 3). This postfocus lowering can be clearly seen in Fig. 4. The curves in Fig. 4 further indicate, however, that there is a sharp drop in f_0 even in the unstressed syllable of the focused word following the stressed syllable. Fig. 5 shows the mean f_0 in semitone at different locations in the poststress syllable broken down by focus and lexical stress. Only word-1 and word-3 sentences are included, because in word-5 sentences, "niece" does not have any poststress syllable. As can be seen in Fig. 5, the downward slope is shallower when the poststress syllable is unstressed than when it is stressed. At the same time, poststress f_0 drop is faster when the stressed syllable is

Table 2
Mean values of various measurements in the post- and prestress syllables under the effects of focus and position, together with probability values from 2-factor repeated measures ANOVAs

	Focus		Position			
	Yes	No	Word 2	Word 3	Word 4	Word 5
Maxf$_0$-post-word 1 (st)	6.2	7.4	8.9	6.5	5.9	6.0
	$F(1,6) = 29.62$,	$\boldsymbol{p<0.01}$	$F(3,18) = 23.49$,		$\boldsymbol{p<0.0001}$	
Maxf$_0$-post-word 3 (st)	6.5	6.8			7.2	6.1
	$F(1,6) = 9.75$,	$P<0.05$	$F(1,6) = 7.88$,		$P<0.05$	
	Focus		Position			
	Yes	No	Word 1	Word 2	Word 3	Word 4
Maxf$_0$-pre-word 3 (st)	9.3	9.9	9.9	9.3		
	$F(1,6) = 3.17$,	*NS*	$F(1,6) = 19.97$,		$\boldsymbol{p<0.01}$	
Maxf$_0$-pre-word 5 (st)	8.6	8.6	10.0	9.5	7.8	7.2
	$F(1,6) = 0.03$,	*NS*	$F(3,18) = 82.86$,		$\boldsymbol{p<0.0001}$	

p values smaller than 0.0125 (after Bonferroni adjustment for multiple comparisons) are printed in boldface.

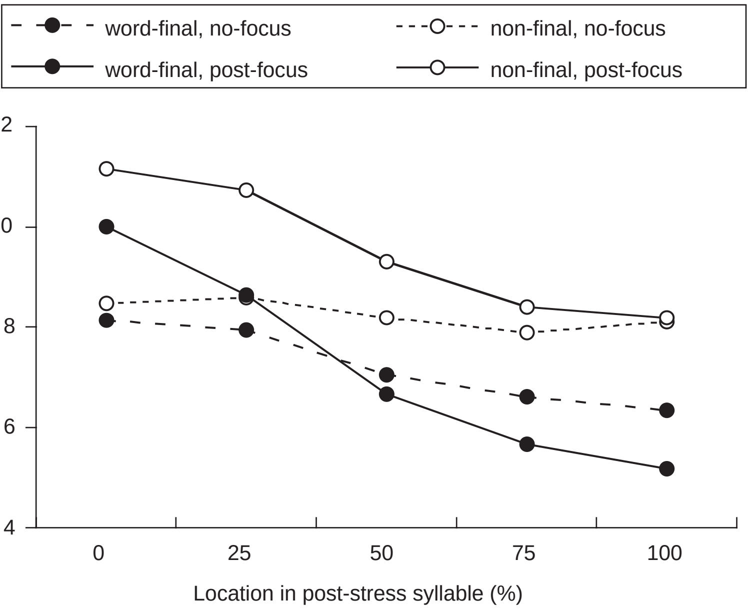


Fig. 5. Mean f_0 in semitones at different locations in the poststress syllable broken down by focus and stress location.

focused than when it is not focused, whether or not the poststress syllable is an unstressed syllable within the focused word. A 4-factor (focus, lexical stress, word position and location in syllable) repeated measures ANOVA finds a significant interaction between focus and word position [$F(1,6) = 11.22$, $p<0.05$], confirming that f_0 drops sharply within the poststress syllable. (The effect of focus is nonsignificant, but those of lexical stress and word position highly significant [$F(1,6) = 61.65$, $p<0.001$; $F(4,24) = 68.64$, $p<0.0001$]. There is also a significant interaction between focus and lexical stress [$F(4,24) = 18.37$, $p<0.0001$]. Hence, the high maximum f_0 of the first postfocus syllable is immediately followed by a sharp fall toward a much lower f_0. And this fall seems to be due to an active lowering of f_0 immediately after the stressed syllable under focus, whether or not the following syllable is part of the focused word.

The lower half of Table 2 shows that maximum f_0 of a word is not significantly different whether or not it precedes a focus. This is despite the fact that for some subjects the target syllables seem to have lower f_0 maxima when they are prefocus than when there is no focus in the sentence, as can be seen in Fig. 4.

Another question that needs to be answered is whether postfocus words are totally devoid of independent local f_0 movements. Fig. 6 shows percentage of discernable postfocus f_0 peaks and the rise size of these peaks in semitone in sentences with initial focus (in sentence group 1 listed in Section 2.1). A peak is discernable if there is an f_0 point between the onset and offset of the words "know" and "niece" (which typically carry F_0 peaks in the neutral focus condition) that is higher than both the starting and ending f_0 of the word. The graph on the left indicates that there are a greater number of discernable peaks when there is no narrow focus than when either word 1 or word 3 is focused. A 2-factor repeated measures ANOVA with *focus* and word *position* as independent variables finds the effect of focus highly significant [$F(1,6) = 83.68$, $p<0.0001$], but

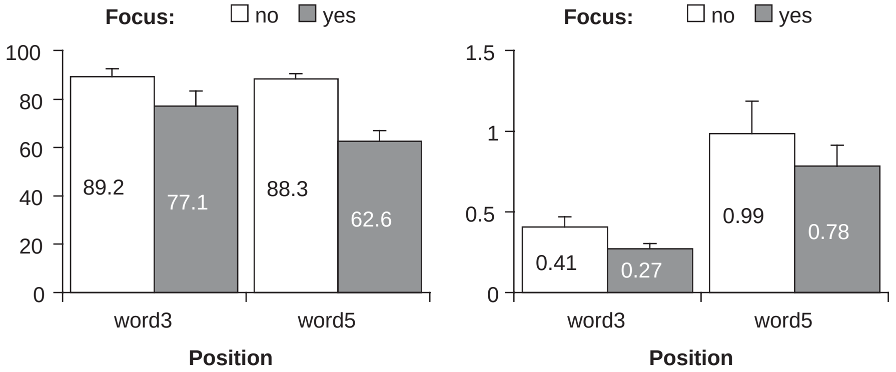


Fig. 6. Percentage of discernable postfocus f_0 peaks (left) and size of the peaks (right) in the postfocus stressed syllables. A peak is discernable if there is an f_0 point between the onset and offset of the words "know" and "niece" that is higher than both the starting and ending f_0 of the word.

the effect of word position nonsignificant. Nevertheless, the lowest percentage of peak occurrence in postfocus condition is still over 60%. The graph on the right in Fig. 6 shows that there is also a difference in rise size between the focus and neutral focus conditions. However, a 3-factor repeated measures ANOVA finds the effect of focus to be nonsignificant, but the effect of word position significant [$F(1,6) = 11.84$, $p < 0.05$]. There is also no significant interaction between focus and word position. Note that although the mean rise size is quite small overall, the rise occurs in a declining f_0 contour. So the size of the intended f_0 movement is actually larger than the observed rise size.

Finally, previous studies have found that sentence-final focus does not produce f_0 patterns different from those in a neutral focus sentence (Cooper et al., 1985). To test whether this is the case for the present data, a 2-factor repeated measures ANOVA was performed on maximum f_0 of word 5. The effect of focus turns out to be significant, with maximum f_0 being higher under final focus (9.3 st) than when there is no narrow focus (6.9 st), $F(1,6) = 14.793$, $p < 0.01$.

To summarize, the effect of a narrow focus is to increase the size of the f_0 peak in the stressed syllable under focus, lower all the postfocus f_0, including that of the poststressed syllables in the focused word, and leave prefocus f_0 largely intact. The general locations of the f_0 peaks are largely the same with or without narrow focus. And the postfocus f_0 lowering does not eliminate the f_0 peaks associated with the stressed syllables. There is initial evidence, however, that the shape of f_0 peak in a word-final stressed syllable does change under focus, especially when the syllable is sentence final.

3.2. Alignment of f_0 peaks and valleys in and around the key words

Through visual inspection, we observed the following patterns and trends in terms of the location of f_0 peaks in and around the key words, most of which can be seen in Fig. 4:

1. In the stressed syllables of all the key words, the f_0 rise starts around syllable onset.

2. If the lexical stress is word-final (“Lee”, “Lamar” or “lure”), the f_0 peak usually occurs before but close to the end of the stressed syllable.
3. If the lexical stress is not word-final (“Nina”, “Emily”, “Ramona”, “mimic” or “minimize”), the peak mostly occurs in the unstressed syllable following the stressed syllable.
4. In the sentence-final monosyllabic word (niece), the peak occurs around the middle of the stressed syllable.
5. f_0 peak occurs earlier when the vowel of the stressed syllable is phonologically (and phonetically) long (“Lee”, “Lamar”, “Nina”, “Ramona”, “lure”, “nanny”) than when the vowel is short (“Emily”, “mimic”, “minimize”, “mummy”).

3.2.1. Alignment of f_0 valleys

The visual observation that f_0 rises consistently start from the beginning of the stressed syllable in the key words is largely confirmed by the measurements *C1-to-min*f_0 and *valley location* (cf. Section 2.5 for definition). Two 2-factor repeated measures ANOVAs were performed to assess the effects of focus and word position on the two measurements. The results are shown in Table 3. The largest mean value of C1-to-minf_0 is 16.9 ms when there is no narrow focus. But even this value corresponds to only 6.7% of the duration of the stressed syllable. Neither focus nor word position has a significant effect on the two measurements.

3.2.2. Alignment of f_0 peaks

The visual observations listed earlier indicate that peak location is potentially related to four factors: focus, word position, lexical stress, and phonological length of stressed vowel (vowel length). The last two factors, however, are not fully independent of each other in the data set. Their effects therefore have to be examined separately. Also, because a word-final open syllable with a phonologically short vowel cannot bear lexical stress, we excluded words with short stressed syllables (“Emily”, “mummy”) when examining the effect of stress location within word, and excluded words with final stress when examining the effect of vowel length. And, because lexical stress and vowel length fully coincide at the word 3 position, this position is not included in the alignment analysis reported next. The alignment patterns in those words, nevertheless, did conform to the same pattern as in the other two positions. Two separate sets of 3-factor repeated

Table 3
Mean values of C1-to-minf_0 and valley location (= 100 × C1-to-minf_0/stress-dur) under the effects of focus and position, together with probability values from 2-factor repeated measures ANOVAs

	Focus		Position		
	Yes	No	Word 1	Word 3	Word 5
C1-to-minf_0 (ms)	3.9	16.9	14.3	7.9	8.9
	$F(1,6)=2.10$,	*NS*	$F(2,12)=0.44$,		*NS*
Valley location (%)	1.2	6.7	5.8	3.4	2.7
	$F(1,6)=2.21$,	*NS*	$F(2,12)=0.47$,		*NS*

None of the effects are significant.

Table 4
Mean values of peak location (= 100 × C1-to-max/syllable-dur) under the effects of focus, lexical stress (upper half), vowel length (lower half) and word position, together with probability values from 3-factor repeated measures ANOVAs

	Focus		Lexical stress		Position	
	Yes	No	Final	Nonfinal	Word 1	Word 5
Peak location (%)	78.5	81.2	68.9	90.8	97.2	62.5
	$F(1,6) = 0.233$	*NS*	$F(1,6) = 60.83$	$\boldsymbol{p < 0.001}$	$F(1,6) = 35.32$	$\boldsymbol{p = 0.01}$
	Focus		Vowel length		Position	
	Yes	No	Long	Short	Word 1	Word 5
Peak location (%)	83.6	124.1	87.6	120.1	130.6	77.1
	$F(1,6) = 20.92$	$\boldsymbol{p < 0.01}$	$F(1,6) = 86.75$	$\boldsymbol{p < 0.0001}$	$F(1,6) = 50.91$	$\boldsymbol{p = 0.001}$

p values smaller than 0.025 (after Bonferroni adjustment for multiple comparisons) are in boldface.

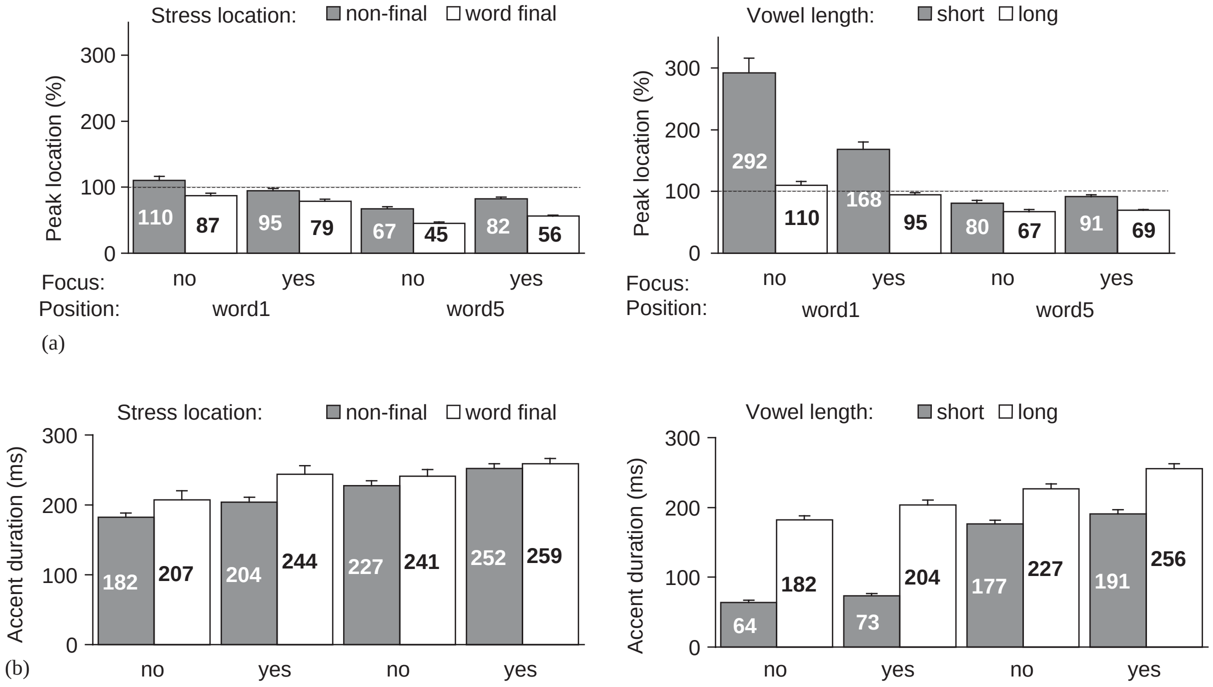


Fig. 7. (a) Mean values of peak location broken down by focus, word position, lexical stress (left) and vowel length (right). Bars higher than the 100% horizontal line indicate f_0 peaks after the stressed syllable. (b) Mean duration of the stressed syllable in word 1 and word 5 according to focus and lexical stress (left), and focus and vowel length (right).

measures ANOVAs were performed and the probability values together with the means are displayed in the upper and lower halves of Table 4, respectively. Also shown in Table 4 are mean values of peak location broken down by focus, word position, lexical stress and vowel length.

From Table 4 we can see that the effect of focus is significant when peak location is broken down by vowel length, but not when broken down by lexical stress. Peak location is earlier when under focus, although the differences are not always statistically significant. In contrast to focus, lexical stress, vowel length and word position all have highly significant effects on peak location. Fig. 7a shows the mean values of peak location broken down by word position, lexical stress and vowel length. In the figure, we can see that f_0 peaks tend to occur early under three conditions: when lexical stress is word-final, when vowel length is long, and when the word is sentence final.

We saw earlier in Table 2 that the duration of the stressed syllable decreases with word position in sentence in an orderly manner: word 1 < word 3 < word 5. This agrees with the trend in the right panel of Fig. 7a quite well. To verify the possibility that it is the shortened syllable duration that pushes the f_0 peak location rightward, we recomputed mean duration of the stressed syllable in word 1 and word 5 according to focus, lexical stress and vowel length. They are displayed in Fig. 7b. In the graph a general trend can be seen: the longer the duration of the stressed syllable, the earlier the location of the f_0 peak. In general, it is when the duration of the stressed syllable is shorter than 200 ms that the f_0 peak occurs in the following syllable, with the exception of the sentence final position. It seems that f_0 peaks tend to occur earlier in the sentence final position, especially when the stressed syllable is sentence-final.

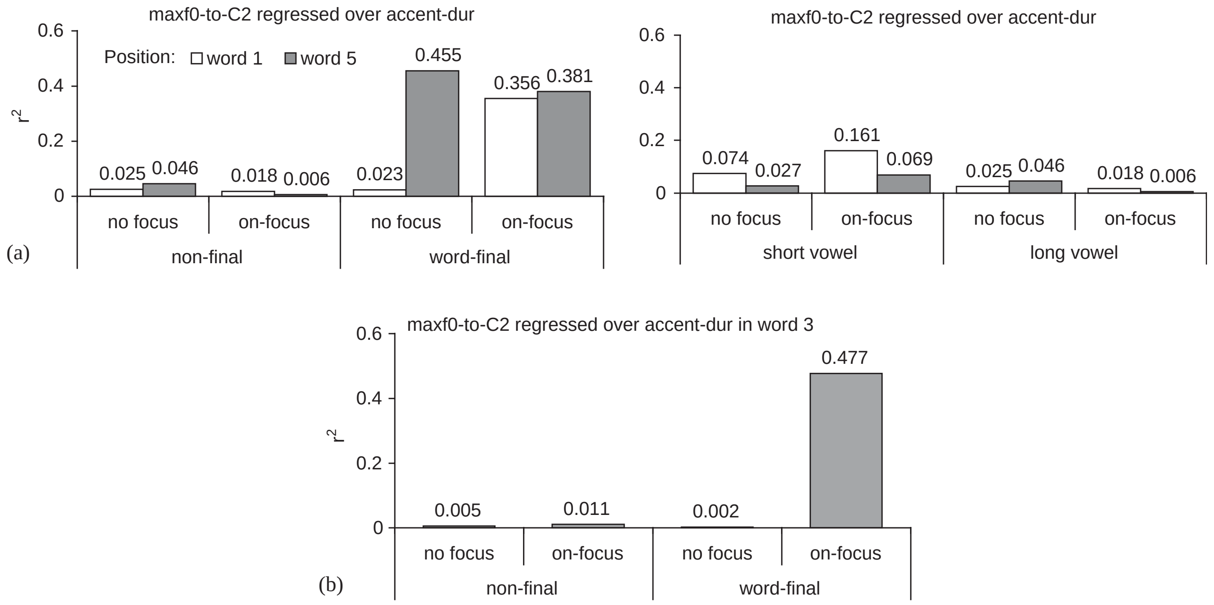


Fig. 8. (a) Results of regression analyses on word 1 and word 5 with stress duration as predictor and maxf$_0$-to-C2 as dependent variable. Left: results broken down by focus and lexical stress. Right: results broken down by focus and length of stressed vowel. (b) Results of regression analyses on word 3 with stress duration as predictor and maxf$_0$-to-C2 as dependent variable.

The visual observations listed earlier indicate that peak location is related both to the location of the stressed syllable in word (whether word-final or not) and to the phonological length of the vowel in the stressed syllable (short or long). To determine whether location of the stressed syllable in word or vowel length is dominant in determining f_0 peak alignment, a set of simple linear regressions were performed using duration of stressed syllable as predictor and maxf$_0$-to-C2 as dependent variable. Maxf$_0$-to-C2 measures f_0 peak delay beyond the offset of the stressed syllable. Fig. 8a displays the r^2 values for word 1 and word 5. In the left-hand graphs, the results are broken down by focus and lexical stress. (Words with short stressed vowels are excluded as explained earlier.) As can be seen, when the stress is word-final (“Lee”, “Lamar” and “niece”) and on-focus, the r^2 values are quite large. When the sentences have no narrow focus, the r^2 values are all very small except when the stress is sentence final (“niece”). The right-hand graph of Fig. 8a shows the regression results broken down by focus and vowel length. (Words with final stress are excluded as explained earlier.) The only r^2 value greater than 0.15 is that of short vowel under focus (“Emily”). A check of the slope of the regression indicates that the amount of peak delay is slightly reduced under focus as the syllable duration becomes longer. All the other r^2 values are very small, indicating that, when lexical stress is not word-final, the location of f_0 peak relative to syllable offset does not change with duration of the stressed syllable. And from Figs. 4 and 7a we can see that the f_0 peaks stay close to the offset of the stressed syllable whether or not the words are under focus.

Fig. 8b displays regression results for word 3. The only sizeable r^2 value for word 3 (sentence-medial position) is that of the word-final syllable (which is also a syllable with long-vowel): $r^2 = 0.477$. This indicates that it is only when the stress is word-final and/or the stressed vowel is long and when the word is under focus that the f_0 peak is affected by duration of the stressed syllable. When the stressed vowel is short and non-word-final, the mean values of maxf$_0$-to-C2 are negative whether on focus or not: -29 and -18 ms, indicating that the peak mostly occurs after the offset of the syllable. This is in contrast to word 5, where maxf$_0$-to-C2 is mostly positive both when the stressed syllable is short and when it is non-word-final.

To summarize, (a) when neither under focus nor sentence final, the f_0 peak associated with a stressed syllable occurs close to and *before* the syllable offset if it is not followed by an unstressed syllable, but close to and *after* the syllable offset if it is followed by an unstressed syllable; in neither case does the peak location relative to syllable offset vary systematically with syllable duration; (b) If the stressed syllable is sentence final or if it is both word-final and under focus, the f_0 peak occurs well before the offset of the stressed syllable and its location becomes increasingly early relative to the syllable offset when the duration of stressed syllable increases. These patterns are rather similar to those reported by Silverman and Pierrehumbert (1990).

3.3. f_0 of unstressed and “unaccented” syllables

The following analyses were performed to help us determine if f_0 contours of “unaccented” syllables are derived through interpolation or target approximation as discussed in the Introduction. Here an “accent” refers to a clearly observable prominent f_0 movement, which, as can be seen in Fig. 4, consistently occurs in or near the stressed syllables of the key words. The goal is to assess the relative influence of the preceding and following “accents” on an

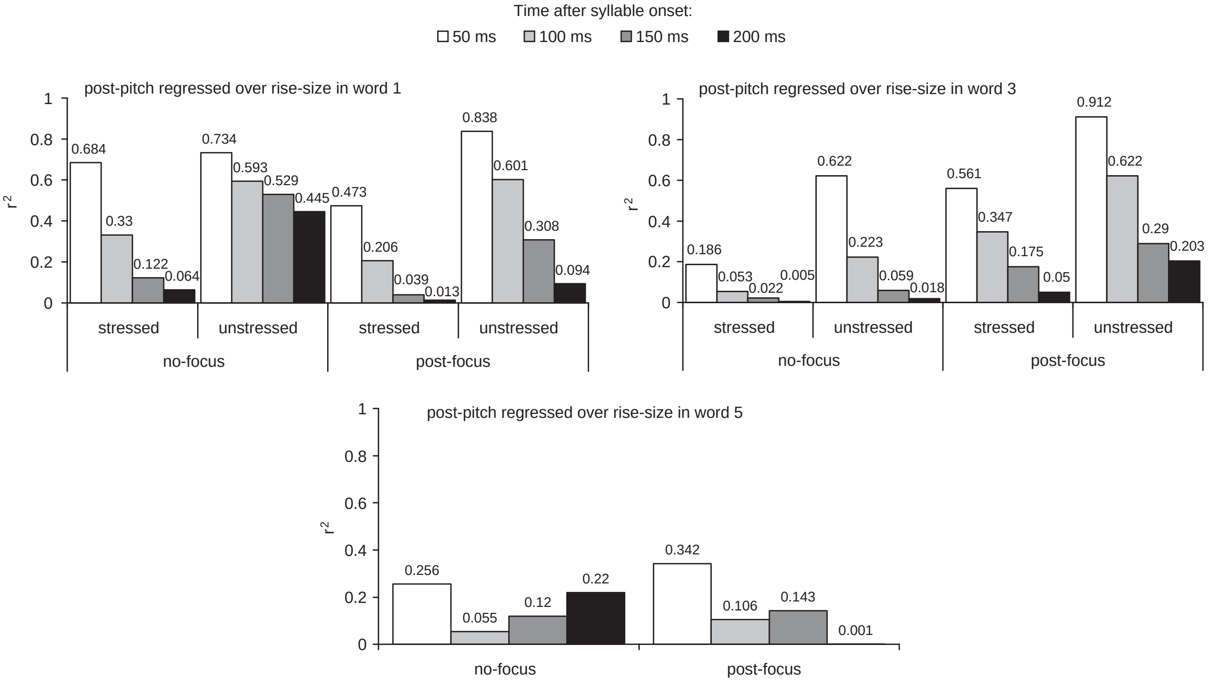


Fig. 9. Results of regression analyses on f_0 height at different locations in the syllable immediately after the stressed syllables in the key words.

Table 5
r^2 values of simple linear regressions on f_0 height at different locations in the syllable immediately preceding the stressed syllable in each word location

Word in sentence	Focus	Stressed?	Location within syllable		
			50 ms	100 ms	Onset
1st	No-focus	Yes	0.013	0.038	0.015
		No	0.03	0.034	0.038
	Prefocus	Yes	0.044	0.022	0.009
		No	0.013	0.081	0.002
3rd	No-focus	Yes	0.0005	0.042	0.348
		No	0.035	0.02	0.285
	Prefocus	Yes	0.124	0.143	0.00002
		No	0.04	0.032	0.001
5th	No-focus	Yes	0.003	0.03	0.086
		No	0.165	0.002	0.065
	Prefocus	Yes	0.037	0.018	0.040
		No	0.00001	0.087	0.005

"unaccented" syllable. We performed several sets of regression analyses on the f_0 height at different locations in the syllables between the stressed syllables in the key words.

Fig. 9 displays the results of simple linear regressions with rise size (in semitone relative to the minimum f_0 of the word) as the regressor and postpitch at 50, 100, 150 and 200 ms after the stressed syllable of the key word as the dependent variables. Postpitch is computed by subtracting the minimum f_0 of the word from the f_0 values at the four locations in the poststress syllable. The r^2 values indicate how much of the variation in postpitch can be accounted for by the height of the preceding "accent" as represented by rise size. As can be seen, postpitch at 50 ms after the poststress syllable can be well predicted by rise size in word 1 and word 3 positions. The prediction is not as good in word 5, although it can still account for 25.6% and 34.2% of the variance for the no-focus and postfocus conditions, respectively. The predictability reduces over time. But the rate of reduction is faster when the poststress syllable is stressed ("may" after "Lee" and "Lamar") than when it is unstressed (in "Nina", "Ramona" and "Emily").

Table 5 displays the results of simple linear regressions with rise size as the regressor, but the dependent variables are prepitch at three points prior to the onset of the stressed syllable of the key word: 50 ms earlier, 100 ms earlier, and at the start of the "pre accent" syllable. Similar to postpitch, prepitch is computed by subtracting minimum f_0 of the word from the f_0 values at three locations in the "pre accent" syllable. As can be seen in Table 5, prepitch is overall poorly predicted by rise size. Only in word 2 are there r^2 values over 0.2, and those are at locations farthest away from the stressed syllable of the key word. Since they occurred only in two conditions in word 2, it is difficult to determine if these higher r^2 values reflect a real anticipatory influence or are merely accidental. Thus there appears to be little evidence for consistent influence of the "accented" syllables on the f_0 of the "pre accent" syllables.

The foregoing analyses demonstrate that the f_0 of a "weak," i.e., unstressed and "unaccented," syllable is extensively influenced by the f_0 of the preceding syllable, but the influence fades away quickly over the course of the weak syllable. Meanwhile, the upcoming syllable has little influence on the f_0 of the weak syllable. This suggests that there exists a *local* destination for the f_0 movement in a weak syllable that is independent of the f_0 of both the preceding and following "strong" syllables. The pitch value of such a destination is best indicated in cases where the following "strong" syllable or word exhibits both extreme high and extreme low f_0 values. In the present data, only the final word of each sentence has such a property, as can be seen in Fig. 5. To estimate the value of the f_0 destination, we measured the offset f_0 of the penultimate word "my",

Table 6
Average offset f_0 of the penultimate word "my" (row 1), maximum f_0 of the final word (row 2), minimum f_0 of the final word (row 3) and mean of maximum and minimum f_0 of the final word (row 4)

	Neutral focus	Postfocus	Finalfocus
End-f_0 (st) of "my"	84.48 (2.61)	83.11 (2.42)	84.40 (2.45)
Max-f_0-word 5	85.52 (2.74)	84.06 (2.56)	87.98 (3.24)
Min-f_0-word 5	82.72 (2.46)	81.08 (2.40)	81.41 (2.35)
Mid-f_0-word 5	84.12 (2.59)	82.57 (2.44)	84.69 (2.67)

Measurements are broken down according to focus conditions as indicated by the column headers. The standard errors are shown in the parentheses.

maximum and minimum f_0 of the final word "niece", "nanny" and "mummy", and average of the maximum and minimum f_0 in the final word. Table 6 displays these values broken down according to focus conditions (neutral, postfocus and final focus) together with their standard errors (in parentheses). We then performed three 2-factor repeated measures ANOVAs, with focus (neutral focus, postfocus and final focus) and f_0 type (offset f_0 of "my", maximum, minimum and mean of maximum and minimum f_0 of final word) as independent variables. The effect of focus (which determines the global f_0 trend around these words) is always significant [$F(2, 12) = 23.00$, 8.38, 20.18, $p<0.0001$, <0.01, <0.001]. The offset f_0 of "my" is significantly lower than maximum f_0 of the final word [$F(1, 12) = 20.37$, $p<0.01$], significantly higher than minimum f_0 of the final word [$F(1, 12) = 8.64$, $p<0.05$], but not significantly different from the mean of maximum and minimum f_0 of the final word [$F(1, 12) = 0.54$, *NS*]. It thus seems that the finally approached f_0 of "my" is half way between the maximum and minimum f_0 of the final word.

3.4. The case of subject 2

Most of the analyses so far have excluded data from subject 2 because of her extensive inter-trial inconsistency in terms of the basic f_0 patterns. As can be seen in Fig. 4, where all the other subjects would have a high f_0 value, subject 2 sometimes has a low f_0 value, and sometimes vice versa. Informal listening to her sentences suggested to us that she might have used different tonal patterns for the key words as well as the nonkey words. Upon closer inspection, we noticed that such alternate f_0 patterns in terms of the location of peaks and valleys occurred in other sentences as well. When using peak location patterns as reference, we can see what is happening when there is no narrow focus in the sentence: the first f_0 peak often occurs much later, mostly in the middle of or later in the word "may" (59/98 of the trials). In contrast, the first f_0 peak always occurs well before or around the onset of "may" for other speakers. More consistently, the f_0 contour in the stressed syllable in word 3 usually assumes a sharp fall toward a valley near the syllable offset (69/98 of the trials), indicating that this speaker actually tried to implement a low pitch for the syllable.

For whatever reason, subject 2 seems to have assigned a low pitch to the second key word in these sentences, and a high pitch to the last key word. This is apparently different from the other 7 subjects examined in the present study. Such free alternation of high and low tonal targets has been suggested before by Goldsmith (1999). Since it occurred in only one subject in the present study, no definitive conclusions can be drawn about it. Also interestingly, the alternation of peak locations by this subject is true only for sentences without a narrow focus on word 1 or word 3. Whenever there is a narrow focus on word 1 or word 3, the location of the f_0 peak becomes quite consistent, and they are not different from the general peak location patterns of other speakers as shown in Fig. 4.

4. General discussion

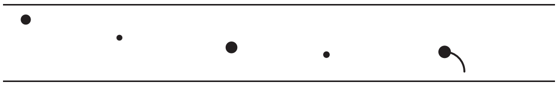
Our analyses of f_0 contours have provided keys to answering the two main questions raised in the Introduction: (1) Is focus realized in parallel or in alternation with other intonational components in English? (2) Are the shape and alignment of f_0 contours in English better accounted for in terms of interpolation between pitch accents or sequential approximation of successive underlying pitch targets in each and every syllable? In the following sections we will first discuss the two questions separately. We will then bring the two together in light of a new model of intonation.

4.1. Manifestation of focus

The gross patterns of focus realization in short declarative sentences in American English are seen quite clearly in Fig. 4. Before offering our interpretation of the results, however, we first attempt to analyze the f_0 curves in Fig. 4 in terms of nuclear tone in the British Tradition (Nucleus) and pitch accents in the AM theory, as shown in the following:

Lee may mimic my niece.

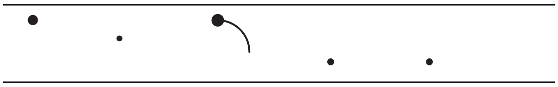
L+H* (L+H*) L+H* L-L% (AM)



(Nucleus)

Lee may MIMIC my niece.

L+H* L+H* L- L% (AM)



(Nucleus)

As we can see, in both systems only the immediately obvious f_0 events are transcribed, while the more subtle ones ignored. In essence, from the perspective of these systems, at any moment in time, either only a single intonational event is happening, or only a single event is worth representing. From a functional perspective as discussed in the Introduction, however, focus is only one of the communicative functions conveyed through f_0. Thus other functions may also be conveyed along with focus. This possibility was investigated in our data analyses designed to answer the two corollary questions specified in the Introduction.

First, as to whether there are local f_0 movements that are independent of focus, both the f_0 plots in Fig. 4 and the postfocus f_0 movement analysis have provided positive answers. Before a narrow focus, the key words have largely the same f_0 peaks as when there is no narrow focus. After a narrow focus, the key words are also associated with small f_0 movements, though with much reduced magnitudes. The analyses of both peak occurrence and size of f_0 rise in postfocus words demonstrate that the percentage of peak occurrence is still over 60% or higher and the size of the postfocus f_0 rises is not significantly different from those in neutral focus sentences (cf. Fig. 6). The existence of postfocus local f_0 peaks agree with focus realization in Mandarin (Jin, 1996; Xu, 1999) as is clearly visible in Fig. 1b–e, in French (Di Cristo & Jankowski, 1999; Delais-Roussarie, Rialland, Doetjes, & Marandin, 2002), and in Neopolitan Italian (D'Imperio, 2001). More interestingly, they also agree with Pierrehumbert's (1980) observation of the so-called "echo-accents" in the area between a nuclear accent and the right edge of the phrase. But Pierrehumbert (1980) treats them as mini-copies of the nuclear accent superimposed on the postfocus region where f_0 is otherwise totally attributable to the phrase accent.[3]

An obvious question now is of course, what are the sources of those off-focus f_0 peaks? A number of functions may be involved. One is lexical stress, which serves a moderate word distinction function in English. The other is a metrical structure that seems to group words into chunks. There is some evidence that in a tone language like Mandarin, such metrical structure is built on alternating strong and weak articulatory strengths (Shih & Sproat, 1992; Kochanski & Shih, 2003). In English, however, the alternation is not only in strength (Liberman & Prince, 1977), but also in local pitch target: high versus nonhigh, as seen in the present data. Yet a third function may have to do with information load of the word (Fowler & Housum, 1987; Hirschberg, 1993; Nooteboom & Kruyt, 1987). Because the present study is not designed to look into these issues, no definitive answer can be clearly drawn from the data. Nevertheless, whatever their causes, the pitch height alternation apparently occur *independently* of focus.

Second, as to whether there are f_0 patterns that are unique to focus, i.e., largely independent of other factors, our data have also provided positive evidence. As just noted, the presence and gross location of the f_0 peaks are not determined by focus. Rather, it is the characteristics of the f_0 peaks that seem directly determined by focus. The data analyses in Section 3.1 show that under a narrow focus, the stressed syllable becomes longer, the maximum f_0 associated with it becomes higher, the size of the f_0 rise becomes larger, and the speed of the rise becomes faster. Furthermore, similar to Mandarin (Xu, 1999), Shanghai (Selkirk & Shen, 1988) and Cantonese (Man, 2002), there are also drastic f_0 changes in syllables after focus in that the maximum f_0 of words following the focused word is much lower than that of the same words in the neutral-focus condition. Note that all of these are *changes* brought by focus to the f_0 peaks that are already there even without a narrow focus. These changes can be most straightforwardly summarized in terms of a three-zone pitch ranges adjustment: expansion under focus, compression after focus, and little or no change before focus. This three-zone pitch range adjustment is therefore what is unique about focus (Xu, Xu, & Sun, 2004).

One direct consequence of the three-zone pitch range manipulation is that immediately following focus there is a sharp f_0 drop. Such a drop has been recognized by both the British tradition and the AM theory. In the former, the entire drop is treated as a high-fall nuclear accent (Crystal, 1969; O'Connor & Arnold, 1961; Cruttenden, 1997). In so doing, the f_0 drop is largely *equated* to focus itself. However, comparing the sentences with nonfinal focus in Fig. 4 with those in Fig. 1, we can see that the f_0 drop around a focused F in Mandarin consists of two parts: that before the syllable offset apparently belongs to F itself (Fig. 1c), because no or little fall is seen inside the H- or R-syllable (Fig. 1b and e), and that after the syllable offset is an apparent transition toward the suppressed pitch range, because it is similar to the sharp F_0 drops after a focused H or R (compare Fig. 1c with b and e).

In the AM theory, the drop is treated as a transition from a H* or LH* to a L- phrase accent that immediately follows (or is spread from the right up to) the accent (Beckman & Pierrehumbert, 1986; Grice, Ladd, & Arvaniti, 2000; Pierrehumbert, 1980). In so doing, focus

[3] We thank one of the reviewers for reminding us of this what we would interpret as a recognition of existence of postfocus accents.

is treated as inherently involving two separate levels of operation: accentual and phrasal. But the fact is that it is the location and scope of focus that *directly* determine the temporal zones of pitch range control. And, as a discourse/pragmatic function, the location and scope of focus are largely independent of both the syntactic and prosodic structures of an utterance (Bolinger, 1972; van Heuven, 1994). In other words, rather than something independently determined by any other factors, the so-called phrase accent is just a description of one of the intrinsic properties of focus.

4.2. Alignment and shape of f_0 contours

The analysis results reported in Section 3.2 reveal the following patterns (which will be referred to as patterns 1–4 in the subsequent discussion):

(1) An f_0 valley always occurs very close to the onset of a stressed syllable whether or not the syllable is focused.
(2) An f_0 peak occurs *near* the offset of a stressed syllable, provided that (a) the syllable duration is about 200 ms or longer, (b) the syllable is not both word-final and on-focus, and (c) the syllable is not sentence final.
(3) An f_0 peak often occurs *after* the offset of a stressed syllable if the syllable duration is much shorter than 200 ms.
(4) An f_0 peak occurs *before* the offset of a stressed syllable that is either (a) both word-final and under focus or (b) sentence final; and the peak becomes increasingly early relative to the syllable offset as syllable duration increases.

One way to interpret these patterns is to assume that these alignments are exactly specified by the intonational phonology of a language. Such an approach is taken by Ladd and colleagues using the AM framework. They argue that f_0 peaks and valleys are "anchored" to the segmental locations exactly as observed (Arvaniti et al., 1998; Atterer & Ladd, 2004; Ladd et al., 1999; Ladd & Schepman, 2003). The AM theory further assumes that the rest of the f_0 contours come from phonetic interpolation between these phonologically specified turning points. The findings about the maximum speed of pitch change (Xu & Sun, 2002) tell us, however, that it takes more than half of the average syllable duration (assuming speech rate of 5–7 syllables/s) to make any noticeable pitch change (e.g., 1–2 st). This implies not only that seemingly long f_0 transitions are inevitably ubiquitous in speech, but also that turning points are intimately related to the transitions. This is because many turning points are in essence the moments in time when one transition ends and the other begins. Thus the consistent valley at the onset of a stressed syllable (pattern 1) may imply that the syllable boundary is where the transition toward a relatively low (or in fact nonhigh, see later discussion) f_0 ends and that toward a high f_0 begins. This understanding is consistent with the intuition of both the British tradition and the AM theory that a pitch accent is associated with a stressed syllable. That is, the implementation of a relatively high pitch coincides, or is synchronized, with that of the stressed syllable, resulting in an f_0 rise that starts at the syllable onset and ends at the syllable offset. Here the f_0 peak is actually the start of the transition toward the nonhigh pitch in the following syllable (pattern 2). If the stressed syllable is short, e.g., in the case of "Emily," "mimic" or "minimize," the rise often ends after the syllable offset (pattern 3). From an articulatory perspective, it also takes time to change the direction of an f_0 movement (Xu & Sun, 2002). Although the implementation of a new, downward f_0 movement may have started at the syllable boundary, the visible shift can take place only when the rise is effectively stopped by the force driving the new movement (cf. Xu, 2001a for a similar account of peak delay in Mandarin). Thus the delayed peak is produced, according to this account, by the same process that produces the peak that is aligned with the offset of the stressed syllable. No phonological peak delay therefore needs to be specified. These patterns are similar to those of Mandarin H, which have either nondelayed f_0 peaks (Xu, 1999), as can be seen in Fig. 1a and b, or delayed f_0 peaks when syllable duration becomes too short (Xu, 2001a).

For pattern 4 mentioned above, again a direct interpretation would be that the peak is simply targeted earlier in a word-final stress under focus. This is the approach taken by Grice et al. (2000) following the tradition of Pierrehumbert (1980). In addition, Grice et al. (2000) treat the *entire fall* after the early peak as a transition from a LH* or H* pitch accent to a L- phrase accent. Similarly, the nuclear tone analysis treats the *entire fall* as a nuclear accent (Cruttenden, 1997). Both theories therefore assume (tacitly or explicitly) that, in these cases, the alignment of the intonational component with the syllable is *readjusted*, so that either the relative location of the entire nuclear accent is shifted earlier in the stressed syllable (British) or the transition toward the L- phrase accent starts well before the offset of the stressed syllable (AM). Note that if the anchor hypothesis (Arvaniti et al., 1998; Atterer & Ladd, 2004; Ladd et al., 1999; Ladd & Schepman, 2003) were to be taken seriously, allowing the readjustment of peak alignment would amount to saying that, for the same level of articulatory operation, the syllable boundary is sometimes fully respected but other times simply ignored.

If, as assumed in our account of patterns 1–3, the implementation of pitch target is synchronized with the syllable, an earlier f_0 peak alignment should be interpreted as an indication that there is a change in the pitch target itself. In fact, an early peak alignment is a characteristic of F in Mandarin, in which the f_0 peak also becomes increasingly early as syllable duration increases (Xu, 1999). Furthermore, as can be seen in Fig. 1b–e, the four Mandarin tones all have rather different f_0 peak alignments relative to the syllable. It is particularly worth noting in Fig. 1b–e how a narrow focus on the first disyllabic word interacts with the lexical tone contours. There is a sharp fall around focus in all plots. But the location of the sharp fall differs across the plots according to the tone of syllable 2. It occurs in syllable 2 when the tone of syllable 2 is L or F (lower plots), but in syllable 3 when the tone of syllable 2 is H or R (upper plots). Assuming that focus realization in Mandarin consists of both on-focus pitch range expansion and postfocus pitch range suppression (Xu, 1999; Xu & Wang 2001; Xu, Xu, & Sun, 2004), the falls occurring at the two locations are rather different in nature. The earlier falls are manifestations of the *tonal* pitch targets themselves (with expanded pitch range); the later falls are the consequences of the postfocus pitch range suppression, i.e., transitions from the high offset f_0 of the preceding H or R to the lowered pitch range, in a similar sense as the transition from a LH* nuclear accent to a L- phrase accent in the AM theory (Beckman & Pierrehumbert, 1986; Grice et al., 2000; Pierrehumbert, 1980). The case of F is especially interesting. Though starting well before the syllable offset, the fall is not completed by the syllable offset. Rather, it continues into syllable 3 until a very low f_0 is reached. This pattern seems comparable to pattern 4 in English: the fall starts before the end of the word-final stressed syllable under focus, but continues into the following postfocus syllable. Therefore, assuming that in English, as reasoned above, local tonal targets are synchronously implemented with syllables, and focus is realized as on-focus pitch range expansion and postfocus pitch range suppression, as discussed in Section 4.1, then the f_0 drop *before* the offset of a word-final stressed syllable under focus is the result of implementing a [fall] target, and the drop *after* the focused syllable would be the result of postfocus pitch range suppression.

As further support for the separation as well as interaction of the local pitch targets and focus, different languages may adopt different strategies in assigning local pitch targets under focus. In Pisa Italian, for example, the peak alignment in a trochaic word seems similar to that of English (patterns 2 and 3) under broad focus (equivalent to our neutral focus) (Gili Fivela, 2002). However, under a contrastive focus (equivalent to our narrow focus), the peak occurs before the offset of the stressed syllable even when not word-final. This alignment pattern contrasts with English where a clear fall occurs within a stressed syllable *only if it is word-final* (present data as well as Silverman & Pierrehumbert, 1990). Thus there seems to be a separation between pitch range adjustments directly due to focus, which may be shared by many languages, and the assignment of local pitch targets to the stressed syllable under focus, which may differ across languages (e.g., English versus Pisa Italian) or even across different lexical tones within the same language (e.g., Mandarin as illustrated in Fig. 1b–e).

4.3. Pitch targets of "weak" syllables

As the results of the regression analyses in Section 3.3 indicate, it is unlikely that the f_0 of "unstressed" syllables comes from interpolation between adjacent accents. Rather, the asymmetrical influence of the tonal context on them seems to support the idea that these syllables are assigned their own pitch targets. The results of the pitch height comparisons show that the finally approached f_0 of the weak syllable "my" is half way between the maximum and minimum f_0 of the final word. These results are consistent with recent findings about the neutral tone in Mandarin (Chen & Xu, 2002, forthcoming). The neutral tone is conventionally assumed to be toneless (Yip, 2002) and the syllable carrying the neutral tone is usually considered to be unstressed (Chao, 1968). Chen and Xu argue that the f_0 of the neutral tone is better understood if it is assumed that (a) the tone has its own static pitch target, which is lower than the target [high] associated with H but higher than the [low] associated with L. They further argue that the target is likely to be [mid], because the final f_0 of a string of neutral tone syllables is half way between the maximum f_0 of a following F and the minimum f_0 of a following L. Based on this close parallel between the Mandarin and English data, we would like to suggest that in both languages each weak syllable is associated with a "default" static target (cf. Yip, 2002 regarding the phonology of a "default" tone) whose pitch value can be represented as [mid].

The regression analyses in Section 3.3 also show that a lexically unstressed syllable is more susceptible than a stressed but "unaccented" syllable to the influence of the preceding "pitch accent." Assuming that in both cases the syllables have their own targets, these differential amounts of carryover influence suggest that there is a difference in terms of the level of articulatory strength applied during the implementation of the targets. This again agrees with the finding of weak articulatory strength in the Neutral tone in Mandarin (Chen & Xu, 2002, forthcoming).

4.4. An alternative model

The foregoing discussion has demonstrated the inadequacies of both the British nuclear tone tradition and the AM theory in accounting for the data obtained in the present study. An alternative model is therefore needed in which the main findings of the present study can fit in naturally. A likely candidate is the Parallel Encoding and Target Approximation (PENTA) model proposed in Xu (2004a, b). The PENTA model defines and organizes the communicative components of speech melody based on function rather than form; it specifies mechanisms for the transmission of *multiple* intonational functions in parallel; and it details *mechanistic* links between the functional components of speech melody and surface f_0 contours. A schematic diagram of the PENTA model is shown in Fig. 10. The stacked boxes on the far left represent individual communicative functions. These functions control f_0 through distinctive *encoding schemes* (shown to their right) that specify the values of the melodic primitives, which include *local pitch target*, *pitch range*, *articulatory strength* and *duration*. The values of the melodic primitives as stipulated by different encoding schemes are specified both symbolically and numerically. Some of the hypothetical symbolic values of the melodic primitives are shown in Table 7.

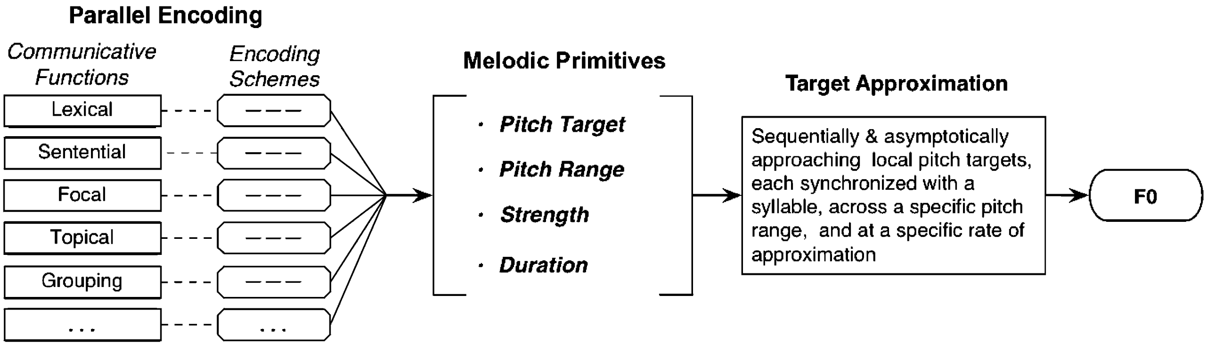


Fig. 10. A brief sketch of the Parallel Encoding and Target Approximation (PENTA) model. See text for explanations. The unnamed block at the bottom left indicates communicative functions yet to be identified.

Table 7
Possible symbolic values of the melodic primitives: *local target*, *pitch range*, *articulatory strength* and *duration*, which may be notationally distinguished from one another by [ ], underline, **boldface** and *italic*, respectively

Local target:	Regular target:	[high], [low], [rise], [fall], [mid]
Pitch range:	Height:	high, low, mid
	Span:	wide, narrow, normal
Articulatory strength:		**strong**, **weak**, **normal**
Duration:		*long*, *short*, *normal*

As can be seen in Table 7, a local pitch target can be either static—[high], [low] or [mid], or dynamic—[rise] or [fall]. When a target is static, its relative pitch height is the only intended goal. When a target is dynamic, both the velocity of the pitch movement and the relative pitch height are the intended goals (Xu & Wang, 2001). Pitch range determines the pitch interval within which local pitch targets are implemented. It has two kinds of specifications: height and span (Ladd, 1996). 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 determines the speed at which a local pitch target is approached. When the strength is **strong**, the target is approached faster than when it is **weak**. Duration specifies the length of the time interval (typically that of syllable) during which a target is approximated.

Corresponding to the symbolic specification of the melodic primitives are also numerical values, which enable them to serve as control parameters of the Target Approximation model (Xu & Wang, 2001) that simulates articulatory implementation of the local targets, as discussed in Section 1.6.3. The process of articulatory implementation of the melodic primitives is one that asymptotically approximates successive local pitch targets, each within the duration of the associated syllable, across a specified pitch range and at a specified speed of approximation. With such a process, the functional components of intonation are realized as continuous f_0 contours through encoding schemes that assign values to the melodic primitives, and articulatory executions that sequentially approximate successive targets.

Applying the PENTA model to English, a hypothetical functional decomposition of the surface f_0 contours of one of the sentences examined in the present study is shown in Fig. 11. Displayed in the graphic part of the figure is the mean f_0 curve of the English sentence "Lee may mimic my niece" said with focus on "mimic" (thick solid line), together with the average f_0 curve of the same sentence with no narrow focus as a reference (thin solid line). The encoding schemes of several likely independent functions are included: Lexical stress, Sentence type (e.g., statement versus question) and Focus. Lexical stress and Sentence type jointly determine local pitch targets; and Focus assigns regional pitch ranges.

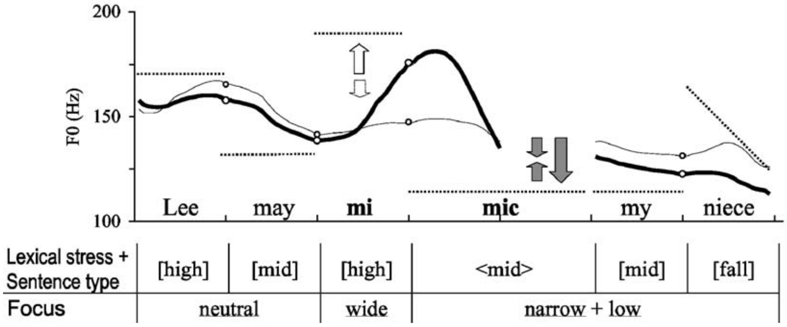


Fig. 11. Decomposition of the f_0 contours of "Lee may mimic my niece" according to the PENTA model. Top: Graphic decomposition. Thick solid curve: focus on "mimic"; thin solid curve: no narrow focus. Short straight lines represent hypothetical local pitch targets. Unfilled block arrows indicate on-focus pitch range expansion. Filled block arrows indicates postfocus pitch range lowering and narrowing. Bottom: symbolic decomposition.

The symbolic representations of the local targets and pitch ranges also correspond to specific numerical values. The specific height and shape of the local pitch targets are depicted by the straight dotted lines in Fig. 11, which can be numerically specified by simple linear functions. The pitch range adjustments by focus are indicated by the block arrows. The unfilled block arrows indicate a wide pitch range (though the downward expansion is not obvious because there happens to be no [low] target in this example). The filled block arrows indicate a narrow + low pitch range.

Note that in Fig. 11, the pitch range expansion is applied only to the stressed syllable in "mimic," while the unstressed second syllable is in the temporal domain of postfocus pitch range narrowing and lowering. This conjecture is based on data discussed in Sections 3.2.2 and 4.2. The apparent f_0 peak near the beginning of the second syllable in "mimic" is likely due to inertia of the larynx whose movement cannot be fully reversed at the syllable boundary. This is because, for one thing, with the average duration of "mi-" being only 168 ms, f_0 is still quickly rising by the syllable offset, and for the other, the following is an unstressed syllable whose weak strength makes it ineffective in reversing the preceding f_0 rise. Such "peak delay" under focus is largely missing in the longer stressed syllables in "Nina", "Ramona" and "nanny," as seen in Fig. 4.

Note also that, the f_0 of "may", "-mic" and "my" is depicted as coming from asymptotic approximation of [mid] targets rather than from interpolation between adjacent f_0 peaks. The height of each [mid], however, is readjusted by the pitch range specification assigned by focus. It could have also been additionally adjusted by other functions not directly controlled in the present study, in particular new topic (or topic shift, turn-taking), which raises the onset pitch of an utterance, resulting in a gradual f_0 decline through the rest of the utterance (Lehiste, 1975; Swerts, 1997; Umeda, 1982). Note that such a decline would have applied not only to unstressed syllables, but also to surrounding stressed syllables.

Finally, in Fig. 11, the f_0 contour in the stressed syllable "mi-" is attributed to a [high] rather than a [rise] target despite the sharp rising across the whole syllable. This is because, firstly, the f_0 peaks mostly occur before the offset of the stressed syllable unless the syllable duration is very short, as in Fig. 11. This is in contrast to R in Mandarin, where the peak always occurs after the end of the syllable, regardless of syllable duration (Xu, 2001a, b). Secondly, the fact that the initial f_0 valleys always occur close to the onset of the stressed syllable, as discussed earlier, contrasts with Mandarin R in which the f_0 valley usually occurs well after the syllable onset and its location becomes increasingly late as the syllable duration increases (Xu, 1998, 1999, 2001a; also see Chen, 2003). Note that by considering the pitch target as a [high], the source of f_0 valley is viewed as belonging to the preceding "unaccented" syllable "may" rather than to the focused syllable "mi-". This is different from Ladd & Schepman's (2003) proposal that the f_0 valley belongs to the accented syllable and the pitch accent should be LH* instead of H*.

In summary, under the PENTA model, communicative components of speech melody are defined and organized based on function rather than form; multiple intonational functions are concurrently transmitted through *encoding schemes* that assign values to the melodic primitives; using these values as control parameters for the Target Approximation model, the functional components of speech melody are ultimately turned into detailed surface f_0 contours. The PENTA model thus provides a framework through which a rich repertoire of communicative functions can be realized concurrently through f_0, with all the details of the f_0 contours still traceable to their proper sources.

4.5. Unresolved issues and future directions

Due to limited scope of the present study, many issues about English intonation are left unaddressed. First, the phenomena of declination and downstep are not dealt with in the present study, because, as argued in Xu (1999, 2001b), they are likely the products of a mixture of factors, including anticipatory dissimilation, carryover assimilation, focus and new topic, among which only focus is systematically controlled in the present study. Only studies with specific designs for addressing those issues can help reveal the mechanisms of downstep and declination.

Regarding the target-syllable synchronization in English, Atterer and Ladd (2004) reported evidence that the exact alignment of f_0 rise onset for similar tonal units differs across languages or even across dialects of the same language. However, the magnitude of such variation is only in the range of tens of milliseconds. So, they concluded that such small differences do not amount to real categorical phonological differences. Furthermore, alignment differences across languages do not necessarily mean asynchrony between syllable and pitch target. They may be reflections of the cross-language gradient differences in the underlying pitch targets themselves rather than gradient differences in the degrees of synchrony across the languages. In this regard we note again that even within a single language (e.g., Mandarin) different alignment patterns can be found across different tonal categories, as seen in Fig. 1 and discussed in Section 4.2. Yet consistent alignments are still found within the same tone (Xu, 1999, 2001a, b). A more definitive way of verifying the synchronization hypothesis for English would be to manipulate the internal structure of syllables while keeping the local pitch targets constant. Xu and Wallace (2004) found initial evidence that the overall f_0 contour alignment with the entire syllable remains constant with variant syllable structures. Nonetheless, further research is needed to determine whether and to what degree pitch targets are synchronized with the syllable in English.

Finally, the findings of the present study are based on General American English. Exactly the same patterns may not be found in other dialects of English. This may especially be true in terms of the local pitch targets associated with either focused or nonfocused syllables.

5. Conclusions

Through examination of detailed f_0 contours in short English declarative sentences with different focus conditions and speaking rates, we found that focus realization in English is fundamentally similar to that in Mandarin, i.e., the pitch range of the focused item is expanded, the pitch range of the postfocus items, if any, is compressed and lowered, and the pitch range of the prefocus items, if any, remains neutral. Such systematic pitch range adjustments generate f_0 contours that have been described by the British nuclear tone tradition and the American AM theory in terms of nuclear tone or nuclear pitch accent combined with low tail or phrase accent. These conventional theories view focus as realized *in alternation* with other f_0-controlling functions. Our data analyses demonstrate, in contrast, that focus is realized *in parallel* with other f_0-controlling functions.

Our analyses also revealed detailed f_0 contours and their alignment with segmental materials that conventional theories have no mechanistic account for. In particular, we found consistent alignment of f_0 valley with the onset of stressed syllable, and consistent alignment of f_0 peak with the offset of stressed syllable when the syllable is non-word-final or word-final but not focused. We also found that f_0 peaks occur well before the syllable offset in word-final stressed syllables that are focused or sentence-final. Neither the British nuclear tone analysis nor the AM theory provides explanations or predictions for these alignment patterns. The only conventional mechanistic account for detailed f_0 contours is provided by the AM theory in regard to f_0 of nonaccented syllables and words, according to which the f_0 values of those syllables come from linear or "sagging" interpolation between surrounding pitch accents (Pierrehumbert, 1980, 1981). Our data analyses demonstrated, however, that f_0 contours between the turning points could not have been derived from interpolation. Instead, both turning points and the intervening trajectories are likely products of a common mechanism, namely, asymptotic approximation of underlying pitch targets that are synchronously implemented with syllables.

Our findings therefore call for an alternative model that defines and organizes the basic intonational components in terms of function rather than form, and specifies articulatorily viable mechanisms for linking those basic components to detailed surface contours. We considered the recently proposed Parallel Encoding and Target Approximation (PENTA) model (Xu, 2004a, b) as a possible candidate. Nevertheless, much remains to be done in future investigation of English intonation, since the present study has only looked into focus in short declarative sentences, and focus is but one of many functions that are conveyed through intonation. Without systematic investigation into each of these functions, our understanding of intonation will remain incomplete.

Acknowledgement

This work is supported in part by NIH Grant DC03902. Our thanks go to Ignatius Mattingly, Fang Liu, Yiya Chen, and Klaus Kohler who read earlier versions of this paper (including a much longer one) and provided insightful suggestions. We are also grateful to three anonymous reviewers who provided valuable comments which helped us make the paper much more focused than the earlier version.

References

Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics*, *36*, 3–25.

Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of f_0: Evidence from German. *Journal of Phonetics*, *32*, 177–197.

Bai, D. (1934). Guanzhong shengdiao shiyan lu [Experiments with tones of Guanzhong dialects]. In *In Shiyusuo Jikan [A Collection by Shiyusuo]* (pp. 355–361).

Beckman, M. E. (1995). Local shapes and global trends. *Proceedings of the 13th international congress of phonetic sciences*, Stockholm (pp. 100–107).

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 255–309.

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

Bolinger, D. L. (1972). Accent is predictable (if you're a mind reader). *Language*, *48*, 633–644.

Chao, Y. R. (1930). A system of "tone letters". *Le Maître Phonétique*, *45*, 24–27.

Chao, Y. R. (1956). Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese. In M. Halle (Ed.), *For Roman Jakobson* (pp. 52–59). Mouton: The Hague.

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

Chen, Y. (in press). Durational adjustment under contrastive focus in standard Chinese. *Journal of Phonetics*.

Chen, Y., & Xu, Y. (2002). Pitch target of Mandarin neutral tone. Presented at *LabPhon 8*, New Haven, CT.

Chen, Y., & Xu, Y. (forthcoming). Production of weak elements in speech—evidence from f_0 patterns of neutral tone in standard Chinese.

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.

Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge: Cambridge University Press.

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

Delais-Roussarie, E., Rialland, A., Doetjes, J., & Marandin, J. M. (2002). The prosody of postfocus sequences in French. In *Proceedings of the first international conference on speech prosody*, Aix-en-Provence, France (pp. 239–242).

Di Cristo, A., & Jankowski, J. (1999). Prosodic organisation and phrasing after focus in French. In *Proceedings of the 14th international congress of phonetic sciences*, vol. 2, San Francisco (pp. 1565–1568).

D'Imperio, M. (2001). Focus and tonal structure in Neapolitan Italian. *Speech Communication*, *33*, 339–356.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, *26*, 489–504.

Gårding, E. (1987). Speech act and tonal pattern in Standard Chinese. *Phonetica*, *44*, 13–29.

Gili Fivela, B. (2002). Tonal alignment in two Pisa Italian peak accents. In *Proceedings of the first international conference on speech prosody*, Aix-en-Provence, France (pp. 339–342).

Goldsmith, J. A. (1999). Dealing with prosody in a text-to-speech system. *International Journal of Speech Technology*, *3*, 51–63.

Grice, M., Ladd, D. R., & Arvaniti, A. (2000). On the place of phrase accents in intonational phonology. *Phonology*, *17*, 143–185.

Gussenhoven, C. (1985). Two views of accent: A reply. *Journal of Linguistics*, *21*, 125–138.

Gussenhoven, C. (in press). Types of focus in English. In: D. Bring, M. Gordon, & C. Lee (Eds.), *Topic and focus: Intonation and meaning. Theoretical and crosslinguistic perspectives*. Dordrecht: Kluwer.

Heldner, M., & Strangert, E. (2001). Temporal effects of focus in Swedish. *Journal of Phonetics*, *29*, 329–361.

Hirschberg, J. (1993). Pitch accent in context: Predicting prominence from text. *Artificial Intelligence*, *63*, 305–340.

Janse, E. (2003). Word perception in fast speech: Artificially time-compressed vs. naturally produced fast speech. *Speech Communication*, *42*, 155–173.

Jin, S. (1996). *An acoustic study of sentence stress in Mandarin Chinese*. Ph.D. dissertation, The Ohio State University.

Kochanski, G., & Shih, C. (2003). Prosody modeling with soft templates. *Speech Communication*, *39*, 311–352.

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

Ladd, D. R., Faulkner, D., Faulkner, H., & Schepman, A. (1999). Constant "segmental anchoring" of f_0 movements under changes in speech rate. *Journal of Acoustical Society of America*, *106*, 1543–1554.

Ladd, D. R., Mennen, I., & Schepman, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America*, *107*, 2685–2696.

Ladd, D. R., & Schepman, A. (2003). Sagging transitions between high pitch accents in English: Experimental evidence. *Journal of Phonetics*, *31*, 81–112.

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

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*, 249–336.

Liu, F. (1924). *Sisheng Shiyan Lu [Experiments with tones]*. Shanghai: Qunyi Shushe.

Liu, F., & Xu, Y. (2003). Underlying targets of initial glides—Evidence from focus-related f_0 alignments in English. In *Proceedings of the 15th international congress of phonetic sciences*, Barcelona (pp. 1887–1890).

Man, V. C. H. (2002). Focus effects on Cantonese tones: An acoustic study. In *Proceedings of the first international conference on speech prosody*, Aix-en-Provence, France (pp. 467–470).

Nooteboom, S. G., & Kruyt, J. G. (1987). Accents, focus distribution, and the perceived distribution of given and new information: An experiment. *Journal of the Acoustical Society of America*, *82*, 1512–1524.

O'Connor, J. D., & Arnold, G. F. (1961). *Intonation of colloquial English*. London: Longmans.

Palmer, H. E. (1922). *English intonation, with systematic exercises*. Cambridge: Heffer.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Ph.D. dissertation, Massachusetts Institute of Technology.

Pierrehumbert, J. (1981). Synthesizing intonation. *Journal of the Acoustical Society of America*, *70*, 985–995.

Pierrehumbert, J. (2000). Tonal elements and their alignment. In M. Horne (Ed.), *Prosody: theory and experiment* (pp. 11–36). London: Kluwer Academic Publishers.

Selkirk, E., & Shen, T. (1988). Prosodic domains in Shanghai Chinese. In S. Inkelas, & D. Zec (Eds.), *The phonology–syntax connection*. CSLI Monograog.

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

Shih, C., & Sproat, R. (1992). Variations of the Mandarin rising tone. In *Proceedings of the IRCS workshop on prosody in natural speech No. 92–37*, Philadelphia (pp. 193–200).

Silverman, K. E. A., & Pierrehumbert, J. B. (1990). The timing of prenuclear high accents in English. In J. Kingston, & M. E. Beckman (Eds.), *Papers in laboratory phonology 1—between the grammar and physics of speech* (pp. 72–106). Cambridge: Cambridge University Press.

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

Umeda, N. (1982). "f_0 declination" is situation dependent. *Journal of Phonetics*, *10*, 279–290.

van Heuven, V. J. (1994). What is the smallest prosodic domain? In P. A. Keating (Ed.), *Papers in laboratory phonology* (pp. 76–98). Cambridge: Cambridge University Press.

Wightman, C. W. (2002). ToBI or not ToBI. In *Proceedings of the first international conference on speech prosody*, Aix-en-Provence, France (pp. 25–29).

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 f_0 contours. *Journal of Phonetics*, *27*, 55–105.

Xu, Y. (2001a). Fundamental frequency peak delay in Mandarin. *Phonetica*, *58*, 26–52.

Xu, Y. (2001b). Sources of tonal variations in connected speech. *Journal of Chinese Linguistics*, monograph series #17 (pp. 1–31).

Xu, Y. (2004a). Transmitting tone and intonation simultaneously—the parallel encoding and target approximation (PENTA) model. In *Proceedings of international symposium on tonal aspects of languages: with emphasis on tone languages*, Beijing (pp. 215–220).

Xu, Y. (2004b). The PENTA model of speech melody: Transmitting multiple communicative functions in parallel. Presented at *from sound to sense: 50+ years of discoveries in speech communication*. Cambridge, MA: MIT.

Xu, Y., & Liu, F. (2002). Segmentation of glides with tonal alignment as reference. In *Proceedings of seventh international conference on spoken language processing*, Denver, Colorado (pp. 1093–1096).

Xu, Y., & 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., & Wallace, A. (2004). Multiple effects of consonant manner of articulation and intonation type on f_0 in English. *Journal of the Acoustical Society of America*, *115*(Part 2), 2397.

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

Xu, Y., Xu, C. X., & Sun, X. (2004). On the temporal domain of focus. In *Proceedings of international conference on speech prosody 2004*, Nara, Japan (pp. 81–84).

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.