ELSEVIER

Speech Communication 33 (2001) 319–337

SPEECH COMMUNICATION

www.elsevier.nl/locate/specom

Pitch targets and their realization: Evidence from Mandarin Chinese

Yi Xu [a,*], Q. Emily Wang [b]

[a] *Department of Communication Sciences and Disorders, Speech and Language Pathology, Northwestern University, 2299 North Campus Drive, Evanston, IL 60208, USA*

[b] *Department of Communication Disorders and Sciences, Rush University and Rush-Presbyterian-St. Luke's Medical Center, USA*

Abstract

In this paper, we propose a preliminary framework for accounting for certain surface F_0 variations in speech. The framework consists of definitions for pitch targets and rules of their implementation. Pitch targets are defined as the smallest operable units associated with linguistically functional pitch units, and they are comparable to segmental phones. The implementation rules are based on possible articulatory constraints on the production of surface F_0 contours. Due to these constraints, the implementation of a simple pitch target may result in surface F_0 forms that only partially reflect the underlying pitch targets. We will also discuss possible implications of this framework on our understanding of various observed F_0 patterns, including carryover and anticipatory variations, downstep, declination, and F_0 peak alignment. Finally, we will consider possible interactions between local and non-local pitch targets. © 2001 Elsevier Science B.V. All rights reserved.

Zusammenfassung

In diesem Artikel schlagen wir einen vorläufigen Rahmen vor, der gewisse Variationen in der Oberflächengrundfrequenz in gesprochener Sprache erklärt. Dieser Rahmen besteht aus Definitionen von pitch targets und Regeln zu ihrer Implementierung. Pitch targets werden als die kleinsten wirksamen Einheiten definiert, die mit linguistisch funktionellen Pitcheinheiten assoziiert sind, und sie können mit segmentellen Phonen verglichen werden. Die Implementierungsregeln basieren auf möglichen artikulatorischen Einschränkungen/Begrenzungen auf die Produktion von Oberflächengrundfrequenzkonturen. Wegen jener Einschränkungen/Begrenzungen kann die Implementierung von einzelnen pitch targets in einer Oberflächengrundfrequenz resultieren, die zu grundeliegende pitch targets nur teilweise wiedergibt. Wir diskutieren ebenfalls den möglichen Einfluss dieses Rahmens auf unsere Interpretation/Auffassung von verschiedenen beobachteten F_0 Mustern, die carryover und antizipatorische Variationen, downstep, Deklination und F_0 Gipfelkoordination beinhalten. Letztlich ziehen wir die Möglichkeit einer Interaktion zwischen lokalen und nicht-lokalen pitch targets in Betracht. © 2001 Elsevier Science B.V. All rights reserved.

Résumé

Dans cet article, nous proposons un cadre d'étude pour rendre compte de certaines variations mélodiques dans la parole. Ce cadre consiste à définir des points cibles de F_0 et les régles gouvernant leurs réalisations. Les cibles sont définies comme les plus petites unités opératives associées à des unités fonctionnelles d'un point de vue linguistique; elles sont comparables à des phonèmes segmentaux. Les règles d'implémentation sont fondées sur des contraintes articulatoires

[*] Corresponding author.
E-mail address: xuyi@northwestern.edu (Y. Xu).

0167-6393/01/$ - see front matter © 2001 Elsevier Science B.V. All rights reserved.
PII: S0167-6393(00)00063-7

influant la production des contours mélodiques. En raison de ces contraintes, la réalisation d'une même cible peut résulter en diverses formes de F_0 qui ne reflètent que partiellement les cibles sous-jacentes. Nous discuterons également de l'intérêt de ce cadre d'étude pour la compréhension de certains contours de F_0, incluant les variations liées au recouvrement et à l'anticipation, la descente graduelle, la déclinaison, et l'alignement des pics de F_0. Enfin, nous considèrerons les interactions possibles entre les cibles locales et non-locales.

Keywords: Pitch targets; Tone; Pitch accents; F_0 contour; Peak delay; Articulatory constraint; Downstep; Declination; Topic initiation

1. Introduction

To understand the acoustic manifestation of speech utterances, it is important to distinguish the underlying functional units from their actual realization in production. In studies of segmental units such as consonants and vowels it has long been known that the surface forms of underlying units vary extensively, and that many of these variations are related to articulatory constraints (Lindblom, 1963; Stevens and House, 1963; Öhman, 1966). It should be reasonable to assume that such is also the case with suprasegmental aspects of speech such as tone, pitch accent, and intonation, which are realized mainly as F_0 contours. Under this assumption, variations in surface F_0 contours result not only from the underlying pitch units but also from the articulatory constraints that determine how these units can be implemented. Past studies have considered a number of F_0 variations that are potentially attributable to articulatory constraints, such as vowel intrinsic F_0 (Ladd and Silverman, 1984; Lehiste and Peterson, 1961; Shi and Zhang, 1987; Steele, 1986a; Whalen and Levitt, 1995), F_0 perturbation by initial consonants (Lehiste and Peterson, 1961; Lehiste, 1975; Howie, 1974; Hombert, 1978; Rose, 1988; van Santen and Hirschberg, 1994), and F_0 declination (Collier, 1987; Lieberman, 1967; Maeda, 1976; Ohala, 1990; Titze and Durham, 1987). In tone research, several studies have attributed certain contextual tonal variations to coarticulation (Abramson, 1979; Gandour et al., 1994; Lin and Yan, 1991; Shen, 1990). In our own research on Mandarin tones in recent years, we have found detailed patterns of contextual F_0 variations in Mandarin (Xu, 1994, 1997, 1998, 1999a) that seem to suggest direct links between certain changes in the shape and alignment of F_0 contours and articulatory constraints. In the present paper, we attempt to make these links more explicit by incorporating them into a preliminary framework for accounting for surface F_0 variations in speech. In this framework, we try to account for a number of observed contextual F_0 variations in terms of (a) the basic pitch targets that are the underlying building blocks of pitch contours, (b) the hosts that carry these pitch targets, (c) the constraints the articulatory system imposes on the realization of the pitch targets, and (d) how pitch targets, their hosts, and articulatory constraints interact to generate surface F_0 contours. Although the framework is mainly supported by evidence from Mandarin tones, we feel that it is potentially applicable to the understanding of F_0 contours in speech in general. We will therefore try to state these implications explicitly so that they can be tested by future studies.

Due to space limitations, we will not compare our framework comprehensively to existing theories of F_0 contour generation. Instead, we will focus on presenting the framework and its implications. In the following sections, we will first outline the framework (Section 2). We will then present supporting evidence from our own studies and other related research (Section 3). Finally, we will discuss the implications of the framework on the understanding of certain observed phenomena of F_0 variation (Section 4). We will also explore how local pitch targets may interact with non-local pitch patterns (Section 5).

2. The framework

We will present the framework by first defining the pitch targets (Section 2.1) and then specifying the rules of their implementation (Section 2.2). The possible articulatory constraints underlying the implementation rules will be discussed later in Section 3.3.

2.1. Pitch targets

The essential assumption of our framework is that observed F_0 contours are not linguistic units per se. Rather, they are the surface realizations of linguistically functional units such as tone or pitch accent. Similar to segmental phonemes, tones and pitch accents are also abstract units, underneath which are articulatorially operable units comparable to segmental phones. We refer to these units as pitch targets:

[1] *Pitch targets are the smallest articulatorially operable units associated with linguistically functional pitch units such as tone and pitch accents.*

In many cases, there is one-to-one correspondence between a pitch target and a tone or pitch accent, while in other cases, a tone or pitch accent may consist of more than one pitch target. For example, in Mandarin, all lexical tones produced in context probably each consists of one pitch target. In contrast, an L tone produced in isolation probably consists of two pitch targets: a [low] followed by a [mid] or a [low] followed by a [high].

We further observe that there are probably two basic kinds of pitch targets – static and dynamic:

[2] *A static pitch target has a register specification, such as* [*high*], [*low*] *or* [*mid*]. *A dynamic pitch target has a linear movement specification, such as* [*rise*] *or* [*fall*].

Fig. 1 is a schematic illustration of two hypothetical underlying pitch targets (dashed lines) and their surface realization (solid line) as proposed in the framework. Two different pitch targets are represented in the figure. The one on the right is a static [low]. The one on the left is a dynamic [rise]. The implementation rule that transforms the pitch targets into the F_0 curves (solid line) will be discussed in Section 2.2.

As an example, Mandarin can be considered, under this framework, to have two static pitch targets – [high] and [low], and two dynamic targets – [rise] and [fall]. They are associated with the four lexical tones in Mandarin: H (high), L (low), R (rising) and F (falling), respectively.

2.2. Implementation of pitch targets

The implementation rules presented in this section concern with how a pitch target is implemented once it is assigned to a segmental unit, referred to here as the host. A host may be a mora, a syllable or an even larger unit. We assume that the assignment is language specific. In Mandarin, for example, the syllable is presumably the host for tonal targets. Regardless of what kind of host a pitch target is assigned to, the framework assumes that,

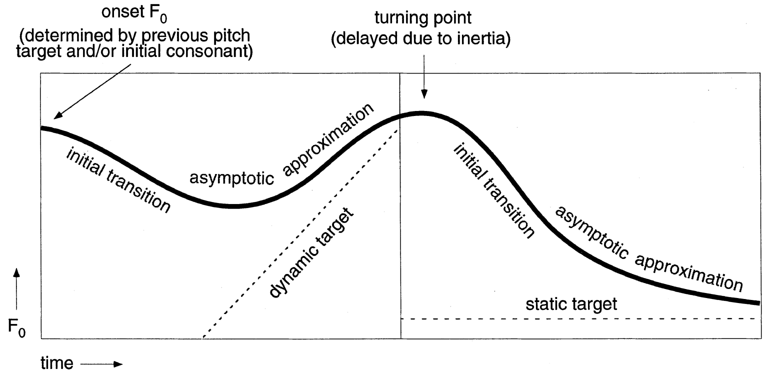

Fig. 1. A schematic illustration of hypothetical pitch targets (dashed lines) and their surface realization (solid curved line). The three vertical lines represent the boundaries of the two consecutive pitch target-carrying units. The level dashed line on the right of the figure represents a static pitch target [low]. The oblique dashed line on the left represents a dynamic pitch target [rise]. In both cases, the targets are asymptotically approximated.

[1] *A pitch target is implemented in synchrony with the host, i.e., starting at its onset and ending at its offset.*

Synchrony between a pitch target and its host, however, does not mean that the F_0 contour in the host always resembles the underlying form of the pitch target. Rather, it only means that the effort to approach the target starts at the onset of the host and ends at the offset of the host. Assuming that the larynx, which implements the pitch targets, cannot change its state instantaneously, approaching a target always takes time, unless the target pitch is already reached before the onset of the host. As a result, as illustrated in Fig. 1, there is often an apparent transition from the initial F_0 at the onset of the host to the F_0 contour later in the host that resembles the pitch target more closely. Furthermore, because the implementation of a pitch target is completed by the offset of the host, the target is realized most adequately in the later portion of the host. In Fig. 1, the steepest F_0 slope for the [rise] occurs in the final portion of the first host, and the lowest F_0 for the [low] occurs in the final portion of the second host.

Fig. 1 also illustrates the manner with which a pitch target is implemented under this framework. That is,

[2] *Throughout the duration of the host, the approximation of the pitch target is continuous and asymptotic.*

Continuous approximation means that the entire F_0 contour during the length of the host is in a continuous process of departing from its initial value, which is dictated by the ending point of the preceding pitch target and the voicing status of the initial consonant (when applicable), and approaching the pitch target proper. This approximation process is not completed until the end of the host.

Asymptotic approximation means that the F_0 movement approaching a target is rapid at first but slowing down gradually over time. By the end of the host, the movement towards the target almost stops as the closest approximation is being achieved.

Worth particular mentioning is the case of dynamic targets. For them, the movements themselves rather than pitch registers are the goal according to pitch target definition [2]. As a dynamic target is gradually approached near the end of the host (see the left half of Fig. 1), the approximation to the rising slope slows down, but the F_0 movement continues. As a result, even after the offset of the host, the rising movement persists for a short while, as is also illustrated in Fig. 1. Presumably, this continued rising movement is due to the fact that *it also takes time to reverse the direction of a pitch movement* (as will be discussed in more detail in Section 3.3.1).

Although we assume that all pitch targets are asymptotically approximated, the rate at which they are approached may vary depending on the direction of the F_0 movement. Specifically,

[3] *A falling F_0 movement is implemented faster than a rising movement.*

This is assumed to be true regardless of whether the fall is due to the implementation of a [fall], or due to a transition from a [high] to a [low].

Finally, the framework assumes that, not only may a pitch target influence the implementation of a following pitch target, it may also influence the implementation of a preceding pitch target:

[4] *A pitch target containing a high pitch point is implemented with a higher F_0 peak when followed by a pitch target containing a low point than when followed by a pitch target with no low point.*

Known as anticipatory raising or regressive H-raising, examples of this phenomenon can be seen in Fig. 3, which will be discussed in detail in Section 3.1.

Note that pitch peaks are not included as targets in this framework. This is because, as we will show later, the occurrence and location of F_0 peaks can be predicted as natural consequences of implementing the static and dynamic pitch targets as defined above.

3. Evidence

Direct evidence supporting the proposed framework comes mainly from our recent findings about contextual tonal variations in Mandarin. The evidence shows how simple underlying tonal targets in Mandarin can generate complex surface F_0 contours, how a dynamic target is inherently different from a static target, and how a pitch target is synchronously produced with its host. Furthermore, literatures on maximum speed of pitch change and on human limb movement control reveal possible underlying mechanisms for the observed patterns of F_0 variation.

3.1. How tonal targets relate to F_0 contours

The four non-neutral lexical tones in Mandarin, known as Tones 1–4, are reported to have typical pitch values of High (55, using the 5-level scale proposed by Chao (1930)), Rising (35), Low (21) and Falling (51), when produced in connected speech in non-final or non-pre-pausal positions. The observed F_0 contours of these tones, however, are much more complex than these pitch values suggest. In particular, the F_0 contour in the early portion of a syllable varies much more than that in the later portion of the syllable (Howie, 1974; Rose, 1988; Shih, 1988). This has led Howie (1974) to conclude that the domain of tone in Mandarin is the rhyme rather than the entire syllable. Similar observations on Zhenhai Chinese have led Rose (1988) to suggest that the perceptually valid F_0 contour should not include the period corresponding to the articulatory transition between two syllables. That transition period, however, is difficult to define, although Rose observes that it should be 50–100 ms after the release of the initial consonant (p. 76). Shih (1988) proposes that different tones in Mandarin have different alignments with the syllable, some starting from the rhyme onset (H and L), but others starting later, even from the middle of the rhyme (R).

Xu (1997) found that a more complete picture about the nature of contextual tonal variations may emerge when the F_0 contours of a tone in different tonal context are systematically compared. In that study, the disyllabic sequence /mama/ was recorded with all combinations of the four lexical tones (not including the neutral tone) in Mandarin produced by eight speakers. Figs. 2 and 3 display F_0 variations in Mandarin H and R due to the preceding tone and the following tone. The graphs in both figures are adapted from Xu (1997). Each curve in the graphs is a (segment-by-segment) time-normalized average of 192 tokens produced by eight speakers.

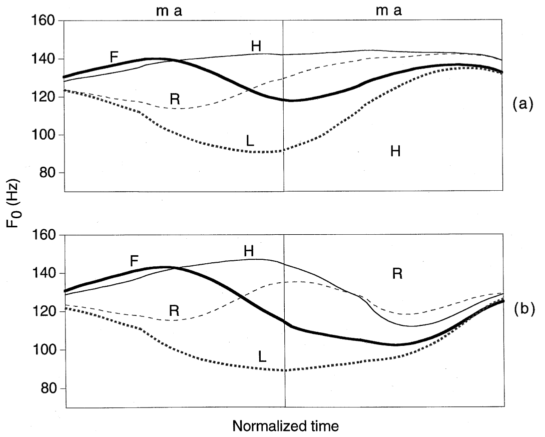

Fig. 2. Effects of preceding tone on the F_0 contour of the following tone in Mandarin. In each panel, the tone of the second syllable is held constant (H in panel a and R in panel b, while the tone of the first syllable varies among H, R, L and F. The vertical lines indicate the syllable boundaries (at the onsets of initial nasals) (Xu, 1997).

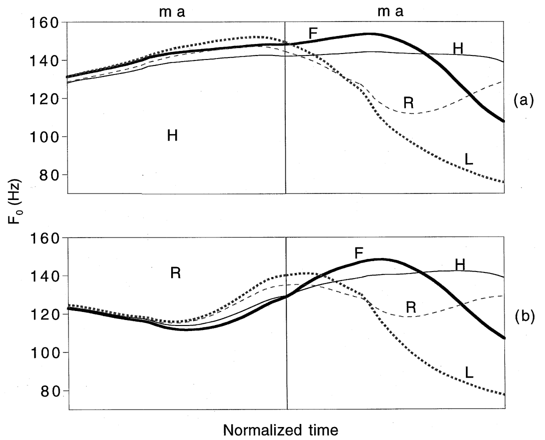

Fig. 3. Effects of following tone on the F_0 contour of the preceding tone in Mandarin. In each panel, the tone of the first syllable is held constant (H in panel a and R in panel b), while the tone of the second syllable varies among H, R, L and F. The vertical lines indicate the syllable boundaries (at the onsets of initial nasals) (Xu, 1997).

In Fig. 2, it can be seen that for any given tone the early portion of the F_0 contour in a syllable always varies with the final F_0 of the preceding syllable, whereas the later portion converges gradually to a contour that seems to conform to the purported underlying pitch values: 55 for H, and 35 for R (Chao, 1968). Due to this gradual convergence, however, the observed F_0 contours are far from optimal in many cases. For H, a considerable portion of the F_0 curve has a rising contour when preceded by F or L, both of which have relatively low offsets. For R, F_0 does not start to rise when following H until well into the vowel. Similar F_0 contour variations due to the preceding tones were also observed in L and F in that study.

In contrast to the substantial carryover variation seen in Fig. 2, the anticipatory variations shown in Fig. 3 is much smaller. Furthermore, the different tones in syllable 2 in Fig. 3 all have very similar F_0 onsets. This seems to be due to the fact that they all share the same preceding tone. Combining the observations on Figs. 2 and 3, we note that the boundary between syllables 1 and 2 serves as an anchor point for both the offset of the preceding tone and the onset of the following tone: the preceding tone continuously approaches its target value up to the syllable boundary, while the following tone starts to depart from that value right at the syllable boundary.

The anticipatory effect shown in Fig. 3 is very different in nature from the carryover effect shown in Fig. 2. The variations are dissimilatory: the F_0 values of a tone are raised when followed by the L tone. Although its underlying mechanisms are still unclear, the anticipatory effect is clear and robust. The phenomenon has been found in a number of languages (Gandour et al. (1994) for Thai; Hyman (1993) for Enginni, Mankon and Kirimi; Laniran (1992) for Yoruba; Laniran and Gerfen (1997) for Igbo; and Xu (1993) for Mandarin). It has been referred to by different names, such as *anticipatory raising* (Xu and Wang, 1997), *regressive H-raising* (Laniran, 1992) and *anticipatory dissimilation* (Gandour et al., 1994; Xu, 1993, 1997).

Further information about the nature of the underlying pitch targets for the tones is revealed by the exact shapes of the F_0 trajectories. In Fig. 2, when the tone of syllable 1 is L, the H and R in syllable 2 have different trajectories. In the case of H, the F_0 curve rises quickly at first, but then slows down gradually before leveling off at a level just below the ideal value (which can be inferred by the ending F_0 values in the HH and RH sequences in the same figure). In other words, the F_0 trajectory of H, when preceded by L, is like a horizontal asymptote approximating a high pitch register: moving rapidly toward the target at first, slowing down as time elapses, and never actually reaching the target even by the end of the syllable. In Fig. 4(a), this kind of horizontal asymptote is schematically illustrated by the filled curve, which continuously approaches the [high] represented by the solid horizontal line.

In contrast, in the LR sequence in Fig. 2, F_0 in syllable 2 starts with a very gradual rise; it picks up speed about three-fifth of the way into the syllable; and it keeps rising at a high speed till the end of the syllable. This kind of F_0 trajectory does not

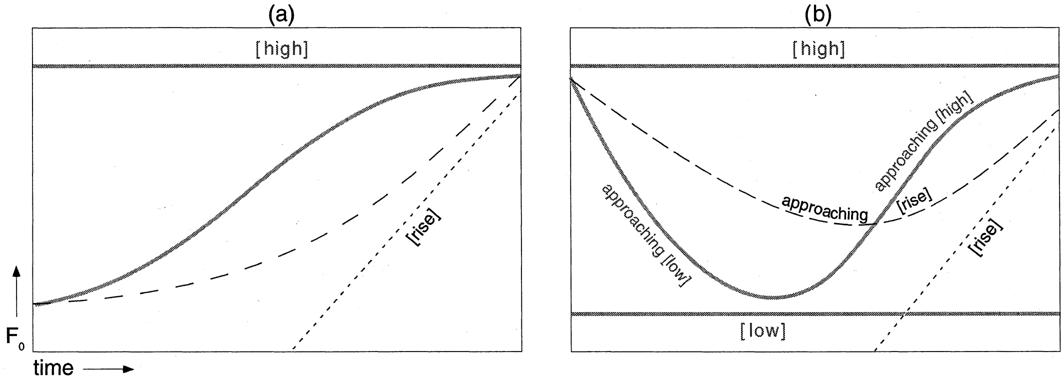

Fig. 4. (a) Hypothetical F_0 trajectories (curved lines) that asymptotically approach a [high] (solid line) or [rise] (dashed line). (b) Hypothetical F_0 trajectories (curved lines) that asymptotically approach two consecutive static targets (solid line) or a single dynamic target (dashed line).

resemble a horizontal asymptote as illustrated by the solid curve in Fig. 4(a). Instead, it appears to be more like a rising asymptote, as illustrated by the dashed curve in Fig. 4(a).

If we assume, however, that R in Mandarin actually consists of two consecutive targets, i.e., a [low] followed by a [high] (Shih, 1988; Wang, 1967), then the asymptotic approximation of the two targets (when the preceding tone ends with a high F_0) should look like the filled curve in Fig. 4(b), which consists of two consecutive asymptotes, the first approaching [low], and the second [high]. As can be seen in Fig. 2(b), this is clearly not the case. Instead, in Fig. 2(b), the F_0 trajectory of syllable 2 in the HR sequence is more like the dashed curve in Fig. 4(b), which is an asymptote approximating a unitary [rise].

There has been a long debate over whether contour tones such as R and F should be considered unitary contour units or sequences of level units such as a low followed by a high and a high followed by a low. While some studies (e.g., Abramson, 1978; Clark, 1978; Pike, 1948; Wang, 1967) argue that contour tones found in languages such as Thai and Mandarin should be considered as single units, others (e.g., Duanmu, 1994; Gandour, 1974; Leben, 1973; Woo, 1969; Yip, 1991) treat them as sequences of high and low targets. The more recent trend seems to favor the sequence account due to the advancement of autosegmental and metrical phonology (Goldsmith, 1979, 1990). In that theoretical framework, because tones are placed on a different tier from the segmental tier, and various rules are devised to associate tones with segmental units, many reported complex tonal phenomena can be represented systematically if all tones are treated as consisting of simple pitch registers. The new evidence that we just discussed, however, seems to indicate that speakers of Mandarin probably produce R and F using unitary dynamic pitch targets rather than sequences of static targets. Further evidence for the existence of dynamic targets will be discussed in the next section.

3.2. How F_0 contours align to syllables

Systematic comparisons of tones produced in different tonal contexts, as demonstrated by Figs. 2 and 3, provide us with information about the nature of F_0 contour variations due to tonal context. They also reveal evidence that syllable boundaries serve as anchor points for the onset and offset of adjacent tones. To observe to what extent syllable boundaries are used as reference points for the alignment of F_0 contours, we can examine the alignment of certain critical points in the F_0 contours, such as peak, valley, the onset of rise and fall, and the location of maximum velocity, with the tone-carrying syllables. In a number of studies, the location of one kind of critical points (F_0 peak) relative to syllable onset were plotted against syllable or rhyme duration, and least-square regression lines best fitting the scattered points were computed (e.g., Arvaniti and Ladd, 1995; Arvaniti et al., 1998; Prieto et al., 1995; Silverman and Pierrehumbert, 1990; Steele, 1986b). In these studies, the relationship between F_0 peak and syllable duration was examined only for their correlation. We note that, since syllable (or rhyme) duration is measured from the onset of a syllable (or rhyme) to the offset of the syllable, and the location of the critical point (F_0 peak) is also measured from the onset of the syllable (or rhyme), plotting the location of the critical points as a function of syllable (or rhyme) duration in fact reveals how these points align with syllable (or rhyme) onset and offset, as illustrated next.

Fig. 5 shows hypothetical regression plots that represent a number of possible patterns of alignment between critical F_0 points and the syllable. The dashed line in Fig. 5 has a slope of 1 and y-intercept of 0. The critical points in cloud 4 in the figure are best fit by this line. This regression function indicates that the critical points in cloud 4 move almost fully in synchrony with the syllable offset: no matter where the syllable offset is, the critical F_0 point always stays close to it. A regression function with a slope of 0.5 as illustrated by cloud 2 indicates that these critical points maintain an equal distance from the syllable onset and offset. A regression function with a slope between 0.5 and 1 as illustrated by cloud 3 indicates that the critical points move more in synchrony with the syllable offset than the onset, whereas a regression function with a slope between 0 and 0.5, as illustrated by cloud 1, indicates that these

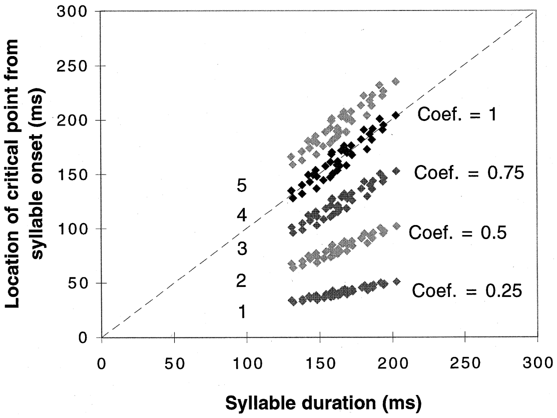

Fig. 5. Hypothetical temporal relations between critical points in F_0 contours and the syllable onset and offset, revealed by plotting the location of the critical point as a function of syllable duration.

critical points move more in synchrony with the syllable onset than the offset. In some cases, the critical points such as F_0 peaks associated with a syllable are delayed beyond the syllable offset. When that happens, the distribution of these points as a function of syllable duration may look like cloud 5 in Fig. 5.

Using the regression analysis paradigm just described, Xu (1998, 1999a) examined the alignment of several critical points in Mandarin, including F_0 peak and valley, and onset of rise and fall (defined as the points of maximum and minimum acceleration, i.e., the second derivative of the F_0 curve). Alignment patterns consistent with several of the clouds in Fig. 5 were found. The first pattern is that the onset of fall in F and the onset of rise in R stay around the center of the host syllable but move in slightly more synchrony with the syllable offset than the onset (i.e., with distributions between clouds 2 and 3 in Fig. 5). The second pattern is that the location of the peak velocity in R always occurs closer to the syllable offset than the onset (i.e., similar to cloud 3) and becomes even more so when syllable duration increases. These two alignment patterns together indicate that, as syllable duration increases, the rising or falling portion shifts more toward syllable offset. This suggests that, when syllable duration varies, what are maintained for R and F are coherent rising and falling contours.

The third alignment pattern is that the F_0 peaks associated with R and the F_0 valleys associated with F usually occur *after* the syllable offset, but nevertheless stay close to and move highly in synchrony with it (i.e., similar to cloud 5). This alignment pattern deserves special attention, because many studies have been concerned with F_0 peak alignment, as will be discussed in Section 4.4. Since an F_0 peak in R or an F_0 valley in F can be observed (and hence measurable) *only* when the ending pitch is different from the starting pitch of the following tone (e.g., in the tone sequences RR, RL, FH and FF), the location of the peak or valley is actually the point where F_0 movement changes its direction. The highly consistent alignment of F_0 peak or valley to the syllable boundary suggests that the boundary between two syllables is the point at which the implementation of one tone ends and that of the next begins. This is another piece of evidence that syllable boundaries are used as reference points for tonal alignment in Mandarin.

The fourth alignment pattern is found in H and L. The F_0 peaks in H and F_0 valleys in L occur before the offset of the host syllable, but nevertheless stay very close to and move much in synchrony with it (i.e., similar to cloud 3). Here again, an F_0 peak or valley is observable only in an appropriate tonal context (e.g., in an LHL or HLH sequence). The fact that the F_0 peaks or valleys occur close to the syllable offset in those contexts indicates that it is near the end of the syllable that the tonal targets are most closely approximated. This suggests that for H and L, just as for R and F, the syllable boundary is used as the reference point for their alignment. On the other hand, the fact that the F_0 peaks and valleys occur before the syllable offset in those contexts indicates that in Mandarin, speakers probably have enough time to reach the pitch targets for H and L, at least at normal speech rate.

Finally, all the aforementioned alignment patterns are found to remain consistent across syllables with different internal structures (i.e., with and without a final nasal), and across different speaking rates (Xu, 1998). This finding suggests that pitch targets in Mandarin tones are aligned to the syllable as a whole with no *direct* reference to its internal structure, further confirming that it is to the syllable that tones in Mandarin are aligned.

Summing up the above discussion, there seem to be two basic types of pitch targets associated with Mandarin tones – static and dynamic. The static pitch targets are specified in terms of pitch registers such as [high] and [low], while the dynamic targets are specified in terms of pitch movements such as [rise] and [fall]. These pitch targets are assigned to the syllable and are produced synchronously with the syllable, as is evident per results of the alignment analyses. The implementation of each pitch target seems to be a process of continuous approximation of the target throughout the tone-carrying syllable. As soon as the syllable boundary is reached, the approximation of the pitch target in the next syllable begins.

3.3. Possible underlying mechanisms

As demonstrated above, simple underlying tonal targets in Mandarin can generate fairly complex surface F_0 contours. To understand why the transitions between tonal targets are so extensive on the one hand, and why tonal targets are so tightly aligned with the host syllables despite the extensive transitions on the other, we need to explore possible underlying mechanisms.

It has long been believed that tones often do not rigidly align with the segmental units (usually syllables) they are associated with lexically. Such belief is based mainly on the observations of phenomena such as tone spreading, tone copying, tone deletion, floating tones, and downstep, etc. These phenomena have been extensively reported for many African tone languages (see, Hyman and Schuh, 1974; Schuh, 1978, for summaries). Our recent investigation into Mandarin tones, however, made two interesting findings. The first is that carryover tonal variations in Mandarin exhibit similar patterns as certain (though not all) tonal variations in African tone languages (Xu, 1997, 1999a). The second is that despite the carryover variations, tones in Mandarin align with their lexically associated syllables very tightly (Xu, 1998, 1999a).

The underlying mechanisms of this dichotomy in tone implementation thus need to be better understood. Two lines of research reported in the literature may provide some insight. The first is the research on the maximum speed of pitch change, which may provide explanations for the widely observed extensive contextual tonal variations. The second line of research is on human limb movement coordination, which, in our view, may shed light on the tonal alignment patterns that we have observed.

3.3.1. Why do surface F_0 patterns vary so much?

Since the F_0 contours associated with tones are produced by the larynx, the mechanical–physiological characteristics of the larynx may impose certain limits that determine the detailed shapes of F_0 contours. The first limit is the maximum speed of pitch change the human larynx is capable of producing. Ohala and Ewan (1973) and Sundberg (1979) asked subjects to change pitch by various amounts as quickly as possible. They then measured the response time, i.e., the amount of time it takes to complete 75% (which is the fastest central portion) of the pitch change. One of their important findings is that the response time is faster for pitch drops than for pitch elevation. Another important finding was that "there was no marked tendency for a change involving a wide pitch interval to take longer than a change involving a smaller interval" (Ohala and Ewan, 1973). To see this more clearly, we computed the maximum rates of pitch change for different pitch intervals using data displayed in Fig. 3 of Sundberg (1979). Fig. 6 plots these pitch change rates as a function of pitch interval width for both pitch rises and falls. If pitch change rate remains constant for different pitch intervals, then the lines in Fig. 6 should be horizontal. Instead, however, the rate of pitch change increases sharply as pitch interval width increases. In other words, although the rate of pitch change can be very fast over a large pitch interval (about 90 semitone/s for a pitch rise of 12 semitones and 124 semitones/s for an equivalent pitch fall), the rates of pitch change over narrower pitch intervals are much slower.

Xu (1999a) found that the average pitch range of Mandarin tones produced without emphasis was about 6 semitones. In the same study, the average duration of non-emphasized syllables was

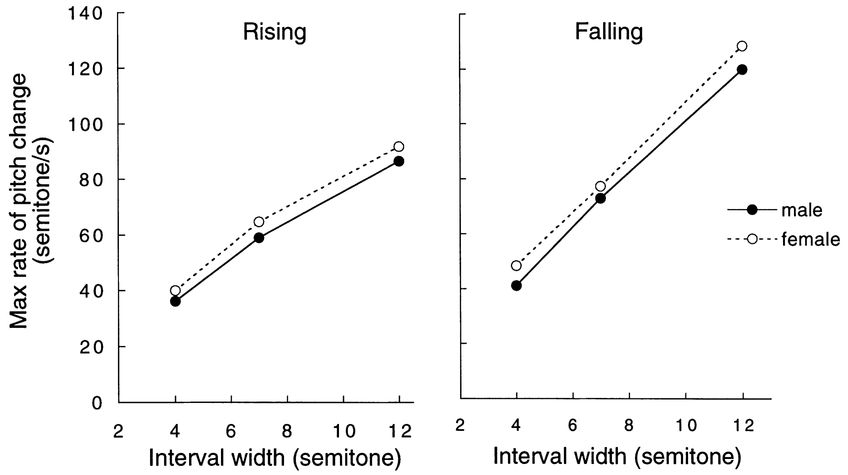

Fig. 6. Maximum rates of pitch change (in semitones per second) plotted as a function of pitch interval width. The pitch change rates are computed from data provided in Fig. 3 of Sundberg (1979).

found to be about 180 ms. Such a duration is probably just enough to allow for two 6-semitone pitch movements to complete, because each should take at least 70–80 ms, according to Ohala and Ewan (1973) and Sundberg (1979). So, when a dynamic tone such as F is produced after L, about half of the syllable duration would be used for the transition from the low F_0 offset of L to the high underlying onset of F, while the other half would be used for the falling contour itself. Similarly, there is often a seemingly long F_0 transition between two tones whenever their underlying pitch values differ substantially at the syllable boundary.

The second articulatory limit on pitch target realization is the maximum speed of pitch direction shift, i.e., the speed at which the larynx can reverse the direction of a pitch movement. As shown in Fig. 3, when R is followed by L, F_0 rises rapidly up to the offset of the R-carrying syllable; it then drops sharply after reaching the peak. To produce such a sharp turn, the larynx needs to first stop the pitch raising gesture and then start a pitch lowering gesture. This process should also take time. Note that this kind of pitch change is different in nature from the pitch change investigated by Ohala and Ewan (1973) and Sundberg (1979). The pitch changes reported in those studies were all in one direction or another, because the subjects were not required to change pitch as many times as possible in a given amount of time (personal communication with Sundberg). While it awaits future studies to actually establish the maximum speed of pitch direction shift, we can at least assume at this point that shifting the direction of pitch movement cannot be instantaneous, and that the time it takes should be considered when trying to understand the shape and alignment of F_0 contours in speech.

3.3.2. Why are tones tightly aligned with host syllables despite the variations?

In speech production, on the one hand, the controls of different acoustic–phonetic aspects of speech sounds seem quite independent of one another. For example, the same vowel can be produced with different fundamental frequencies by the same speaker; and the same pitch pattern (e.g., a tone) can be produced with different vowels. On the other hand, however, it is not known how much freedom a speaker actually has in controlling the relative timing of different articulatory events once they are related at some abstract level. Our finding of rigid tone-syllable alignment in Mandarin (Xu, 1998, 1999a) provides preliminary evidence that such freedom may be quite limited. Given that a tone is lexically assigned to a syllable, speakers consistently produce them in synchrony, regardless of its tonal context.

In searching for possible mechanisms for such synchrony, we take notice of several studies on human limb movement control. Kelso (1984) asked human subjects to perform a simple task of wagging two fingers (one in each hand) together. At low speed, they could start the movement cycles of the two fingers either simultaneously, i.e., with 0° phase shift, or with one finger starting earlier than the other by half a cycle, i.e., with a 180° phase shift. At a high speed, however, they could move the two fingers together only with 0° phase shift, i.e., starting and ending the two movement cycles simultaneously. Schmidt et al. (1990) further found that the same happened when two people were asked to oscillate their legs while watching each other's movement. Based on such findings, these authors suggest that (a) there is a deep-rooted biological tendency to coordinate one's movement with the environment whenever pertinent, regardless of whether the environment is within the same person or between persons, (b) the 0° phase angle is the most stable phase relation between two coordinated movements, and (c) at high speed, the only way to temporally coordinate two movements is to lock their phase angle at 0°, i.e., making them oscillate in full synchrony.

Relating the findings about human limb movement to the synchrony between tone and syllable, we suspect that just as concomitant finger wagging and leg swinging movements are highly coordinated, so are probably the articulatory movements for producing tones and syllables that are lexically associated. From the perspective of movement control, this kind of coordination may in fact greatly reduce the degrees of freedom of the articulatory movements that the human neural system has to control in order to coordinate various articulatory gestures. In particular, the open–close cycles of the syllable production may provide a convenient coordinate structure under which both segments and pitch targets can be organized. Since speech production is a highly skilled activity, fluent speakers may have optimized their articulatory movements so that related gestures are indeed highly coordinated and maximum synergy is achieved.

In summary, three underlying mechanisms may be jointly responsible for relating pitch targets to the surface F_0 contours: (1) the maximum speed of pitch change the larynx can produce, (2) the maximum speed at which the larynx can change the direction of F_0 movement, and (3) articulatory coordination of the production of pitch targets and their hosts.

4. Implications

In the following sections, we will try to demonstrate that our framework may provide explanations for a number of widely reported phenomena related to F_0 contours. Furthermore, we will show that the task of the speaker in pitch production is probably much simpler than the often complicated surface F_0 contours may suggest.

4.1. Carryover F_0 contour variation

First, according to implementation rules [1] and [2] in Section 2.2, a pitch target is implemented in synchrony with its host, and the approximation of each pitch target is continuous and asymptotic throughout the duration of the host. The following pattern of pitch target realization is thus inevitable:

When two pitch targets occur next to each other, if the offset of the first one is different from the onset of the second one, the second one will appear as if it has been assimilated or partially assimilated to the first one.

The surface pattern so predicted seems to resemble certain cases of tone spreading, which are described as assimilation of either a portion of a tone or its entirety to the tone that precedes it (Manfredi, 1993) and have been reported for many tone languages (Hyman and Schuh, 1974). For example, in Yoruba, an H tone is known to spread into the following L tone, changing it into a so-called falling tone; similarly, an L tone is known to spread into the following H tone, changing it into a so-called rising tone (Schuh, 1978). The F_0 tracings presented by Laniran (1992) suggest, however, that there are many similarities between the surface form of tone spreading in Yoruba and the carryover transition in Mandarin as shown in Fig. 2. Since the surface assimilation predicted by our framework is due to the implementation of pitch targets rather than due to a real change of the targets as the spreading analysis may imply, there should be no change of perceived tone categories despite the assimilation. Indeed, as found by Xu (1994), when the F_0 contour of a Mandarin tone is significantly assimilated to the preceding tone, native listeners can still correctly identify the underlying tone as long as the original tonal context is present. It thus awaits perceptual investigations similar to that done for Mandarin to determine if the reported tone spreading causes any real change of tone categories in Yoruba and in other languages known to have similar spreading rules.

4.2. Downstep

Second, because pitch falls are in general faster than pitch rises (implementation rule [3] in Section 2.2), given the same amount of time and effort, a rise cannot cover as much F_0 range as a fall. Therefore, a rise after a fall is unlikely to fully recover the F_0 drop due to the fall, unless extra effort is given to the rise. Furthermore, the effect of anticipatory raising elevates the maximum F_0 of H when it is *followed* by L (implementation rule [4] in Section 2.2). Therefore, if an L tone intervenes between two H tones, the anticipatory raising will make the first H higher while the rate difference between rise and fall will push the second H in the opposite direction. As a result,

In an HLH sequence, the first H is usually higher than the second H in F_0.

This phenomenon is known as *downstep* and has been widely reported for both tone and nontone languages (Clements and Ford, 1979; Hyman and Schuh, 1974; Manfredi, 1993; Pierrehumbert, 1980; Poser, 1984; Prieto et al., 1996; Shih, 1988; Stewart, 1983). As an example, Fig. 7 shows such effect in Mandarin, in which both anticipatory raising and the rate difference between fall and rise are clearly visible.

4.3. Declination

Third, when downstep occurs repeatedly, the F_0 drops it produces can be accumulated. Hence,

A gradual decline in F_0 may result from implementing a sequence of alternating [high] and [low] if there are no other effects to raise F_0 along the way.

This repeated decline should at least partially resemble the phenomena of downdrift and declination, the former reported mainly for tone languages (Anderson, 1978; Hombert, 1974) and the latter for languages in general (Cohen et al., 1982; Ladd, 1984; Maeda, 1976). Interestingly, in the speech synthesis system developed by Mattingly (1966), the declination effect was actually simulated by the rate difference between rise and fall. While the effect of repeated downstep is probably not enough to fully account for downdrift and declination (since other factors such as focus and topic initiation may also contribute, as will be

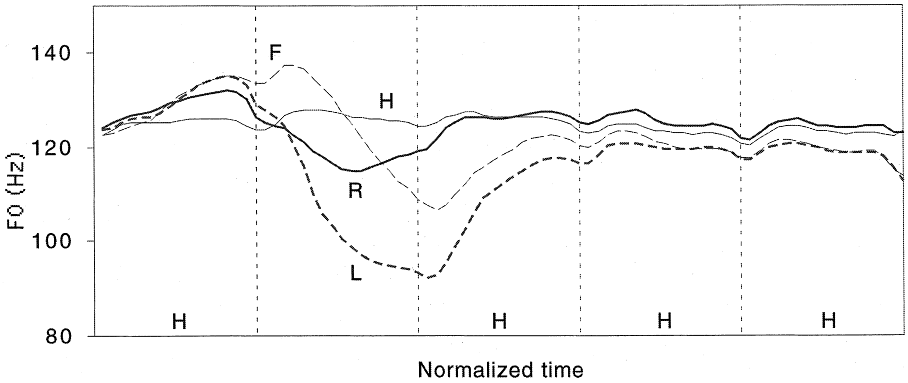

Fig. 7. Average F_0 tracings of sentences consisting of all H tones except on the second syllable which carries H, R, L and F alternatively. Each curve is an average of 20 tokens produced by four male speakers (five repetitions per speaker) (Xu, 1999a).

discussed in Section 5), it is likely to be one of the important contributors.

4.4. F_0 peak location

Finally, although our framework does not include pitch peaks as targets proper, it can nevertheless make predictions about the location of observable F_0 peaks. In general,

The occurrence and location of F_0 peaks may be predicted by (*a*) *the property of the pitch target,* (*b*) *the properties of the adjacent pitch targets, and* (*c*) *the duration of the host.*

First regarding factor (a). Apparently, there has to be a high pitch as part of the target for an F_0 peak to occur. Such is the case in targets like [high], [rise] and [fall]. The property of the pitch target also determines where the F_0 peak is likely to occur. A peak associated with a falling target should occur earlier in the host; a peak associated with a [high] should occur later in the host; and a peak associated with a [rise] should occur just beyond the offset of the host, as shown by Fig. 2 and discussed in Section 3.2.

Factor (b) is the properties of the adjacent pitch targets. An F_0 peak is not likely to occur (or to be easily measurable) when a [rise] is followed by a [high] (in H), as can be seen in the upper panel of Fig. 2. The location of the peak in F may vary depending on the preceding tone, as can be seen in Fig. 8. When preceded by L, it takes a long time for F_0 to reach a sufficient height to make a reasonable approximation of the falling movement, thus delaying the peak.

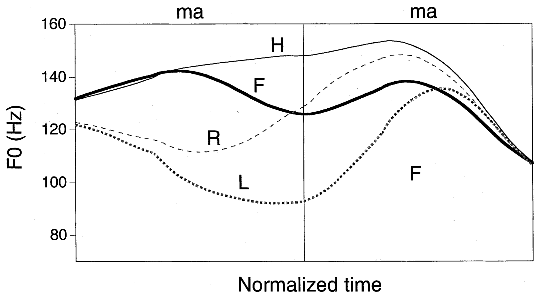

Fig. 8. Effects of preceding tone on F_0 contour of F in Mandarin. The vertical lines indicate the syllable boundaries (at the onsets of initial nasals). Adapted from Xu (1997).

Note also the F_0 peak in the RF sequence in Fig. 8. Without the syllable boundary marking and direct comparison with other tone sequences, it would be difficult to determine the nature of the peak in this sequence. As can be seen in the figure, the last portion of the F_0 contour in syllable 1 remains rising. After the syllable boundary, the F_0 contour continues to rise before reversing its course at the peak. It seems that even if we know that an F_0 contour is composed of a rising–falling sequence, we probably cannot automatically assume that the F_0 peak is where the rising target ends and the falling target starts. For example, Caspers and van Heuven (1993) found that in a rising–falling sequence, the end of the rise varied its alignment depending on whether it is followed by an immediate fall. Silverman and Pierrehumbert (1990) also reported similar variations. As can be seen in Fig. 8, the falling pitch target in F in syllable 2 seems to require a rather high starting point. However, both the location and the height of the F_0 peak in the second syllable varies substantially when preceded by different tones. Apparently, just as in the other three tone sequences in Fig. 8, the F_0 peak in the RF sequence is the result of approximating the [fall] in the second syllable rather than approximating the [rise] in the first syllable.

Factor (c) is the duration of the host that carries the pitch target. When a [high] is surrounded by [low], much of the F_0 contour associated with the [high] will be rising due to asymptotic approximation, as is apparent in the third syllable in Fig. 9. Despite this overall rise, however, the F_0 peak associated with the second H tone in the figure occurs before the syllable offset. As discussed earlier, the mean duration of that syllable is around 180 ms (Xu, 1999a), and that is presumably enough time for the [high] to be reached by the end of the syllable. If, however, the duration of a syllable is significantly shortened, a [high] may not be fully reached within the syllable. In such a case, the F_0 rise may be extended into the early portion of the following syllable. This has been recently confirmed (Xu, 1999b). Further evidence can be found in (Laniran, 1992). In her data, the highest F_0 in an LHL sequence usually occurs in the early portion of the second L-tone syllable. Assuming that it was

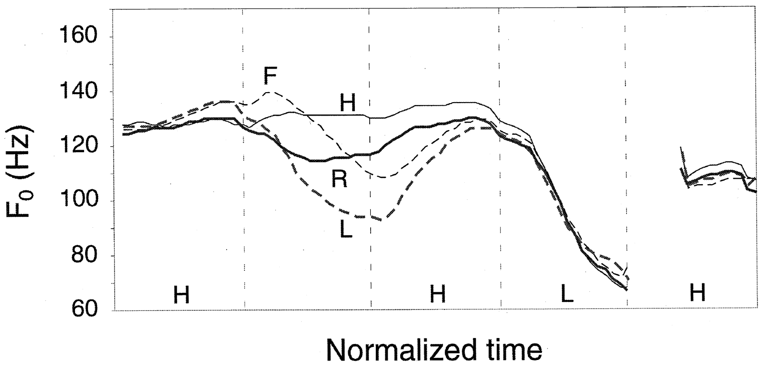

Fig. 9. Average F_0 tracings of sentences consisting of tone sequence of H*t*HLH, where *t* stands for any of the four Mandarin tones. Each curve is an average of 20 tokens produced by four male speakers (five repetitions per speaker) (Xu, 1999a).

due to short syllable duration that this peak delay occurred, then if the tone sequence became LHHL, there should be enough time for the [high] to be reached before the end of the second H-carrying syllable. Indeed, both the F_0 tracings and the two-point-per-syllable measurements presented by Laniran (1992) show that, in most of the L*n*HHL sequences (where *n*H indicates any number of H tones), the highest F_0 no longer occurs at the beginning of the second L-carrying syllable, but inside the last H-carrying syllable.

In summary, our framework therefore predicts that there are two circumstances under which an F_0 peak may be delayed beyond the offset of the host: (a) when the pitch target is [rise] and is followed by a target with a low pitch onset, and (b) when the pitch target is [high] and surrounded by low pitch values and the duration of the host is sufficiently short. Since many studies have reported peak delays (Arvaniti and Ladd, 1995; Arvaniti et al., 1998; Grimm, 1997; de Jong, 1994; Ladd, 1983; Pierrehumbert and Steele, 1989; Prieto et al., 1995; Silverman and Pierrehumbert, 1990; Steele, 1986b), it would be interesting to examine in each case whether the observed peak delay is in any way related to the factors just described.

5. Non-local pitch targets

The pitch targets discussed so far all correspond to the most local pitch units, such as lexical tones. Surface F_0 contours, however, are also affected by many non-local factors. In fact, non-local factors probably specify the *pitch range* over which local pitch targets are implemented. In the following discussion, we consider several such factors to see how they may interact with local pitch targets in generating surface F_0 contours.

The first factor is focus, which has been investigated in many studies (Bruce, 1977; Caspers and van Heuven, 1993; Cooper et al., 1985; Eady and Cooper, 1986; Eady et al., 1986; Gårding, 1987; Jin, 1996; Liberman and Pierrehumbert, 1984; Pierrehumbert, 1980; Prieto et al., 1995; Shih, 1988). To examine in detail the effect of focus on F_0 contours in Mandarin, Xu (1999a) recorded short sentences produced by eight Mandarin speakers. Each of these sentences consisted of three words, the first and last words were disyllabic nouns and the second a monosyllabic verb. The middle three syllables in these sentences varied in tone. The F_0 curve of each utterance was computed vocal-cycle by vocal-cycle and normalized segment by segment. Fig. 10 displays some of the time-normalized F_0 curves averaged over all eight speakers, adapted from Fig. 5 of Xu (1999a). The sentences in each of the two panels have the same tone sequence, but differ in their focus status. Displayed in this way, the effects of focus on local F_0 contours and global F_0 shapes become clearly visible.

As can be seen in Fig. 10, substantial variations occur in the F_0 of each tone as a function of focus. In general, tonal contours under a non-final focus are substantially expanded; those after the focus are severely suppressed (lowered and compressed);

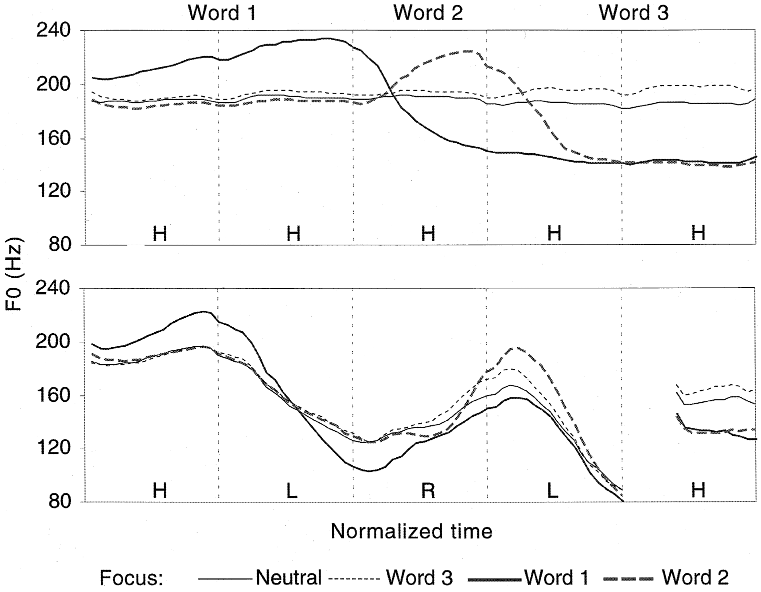

Fig. 10. Focus effects in Mandarin. Each curve is an average of 40 tokens produced by eight speakers. The gap in the last syllable in the lower panel is due to a voiceless stop (Xu, 1999a).

and those before the focus do not deviate much from the neutral-focus condition.

Also from Fig. 10, it seems that each pitch range modification is being implemented continually and asymptotically within its temporal scope. Under focus, the F_0 in the first word in the upper panel rises continually and asymptotically during the two H-carrying syllables. Post-focus, the F_0 in the last two words lowers continually and asymptotically during the three H-carrying syllables: by the end of the second syllable, F_0 virtually levels off. In the lower panel, just as in the upper panel, the post-focus F_0 does not differ whether the focus is on the first or second word.

The pitch range of a local pitch target may also be modified by another factor which may be called *topic initiation*. As found by several studies (Lehiste, 1975; Nakajima and Allen, 1993; Umeda, 1982), the F_0 of the first word or phrase in the first sentence of a paragraph is often much higher than the F_0 in the later portion of the paragraph. Umeda (1982) suggests that an exceedingly high F_0 peak at the onset of the first sentence of a paragraph, which is different from the peaks that occur in stressed syllables in later words, is probably used as a beginning signal for new topics, or so produced to draw listeners' attention. Nakajima and Allen (1993) presented rather convincing data on high F_0 values related to topic-initiation (referred to as topic shift in their study). Although no study we are aware of has looked into it, it is possible that when a declarative sentence is read aloud in isolation, speakers usually say it as if to introduce a new topic. It then may follow that most of the so-called "laboratory speech" probably tend to have a topic initiating F_0 pattern in general.

Assuming that to initiate a new topic a speaker substantially raises the pitch of the very first word (or the first word that can be stressed) in a sentence, assuming further that pitch range after that word drops freely (i.e., without active lowering) to a relatively neutral level, then the overall F_0 level would probably appear to decay exponentially: dropping fast at first, slowing down gradually, and finally virtually leveling off. Interestingly, exponential F_0 decay has been observed by a number of studies: Gelfer et al. (1985) for Dutch, Pierrehumbert (1980) for English, Prieto et al. (1996) for Mexican Spanish, and Shih (1997) for Mandarin. Since these studies were not designed to test the possible contribution of topic initiation to the observed non-linear F_0 decline, it is not known whether the link we just suggested actually exists. This is an area where further investigations are certainly needed.

Based on the above discussion, we propose that there are non-local pitch targets associated with prosodic functions such as focus, topic initiation, and that they are independent of (i.e., orthogonal to) local pitch targets. On this point, our view is similar to the superposition account of intonation (Fujisaki, 1988; Gårding, 1979; Grønnum, 1995; Ladd, 1995).

Similar to local pitch targets, non-local pitch targets are the underlying components of prosodic functions. A prosodic function may consist of more than one non-local pitch target. For example, a non-final focus probably consists of an on-focus target and a post-focus target.

Such non-local pitch targets are presumably pitch range targets in the sense that they specify the pitch range over which local pitch targets are implemented. Furthermore, a pitch range may have two parameters for its specification: (a) its register (e.g., high, mid or low), and (b) its scope (e.g., large or small) (similar to "level" and "span" proposed by Ladd (1996)). For example, the first pitch range target (on-focus) of a non-final focus probably has only one specification: large scope. Its second pitch range target (post-focus), in contrast, probably has two specifications: small scope and low register. The pitch range target for topic initiation, as discussed earlier, probably has only one specification: high register.

Similar to local pitch targets, pitch range targets are implemented asymptotically within its assigned temporal domain. (This domain, however, is not determined by the pitch range target itself.) If the domain is large, then the target can be reached well before the end of the domain, and a relatively stable F_0 over a period of time may result, as can be seen in Fig. 10 when focus is on word 1 (upper panel). If the domain is small, then dynamic patterns similar to that of a local pitch target may result even though the lexical tone under focus is H, as can be seen in Fig. 10 when focus is on word 2 (also upper panel).

Finally, as suggested by various non-experimental studies of intonation (e.g., O'Connor and Arnold, 1961; Chao, 1968; Crystal, 1969; Bolinger, 1989, to mention just a few), there are many more intonation patterns than we have discussed. Few of the intonation patterns described in these studies, however, have been investigated experimentally (but see Hirschberg and Ward, 1992; Pierrehumbert and Hirschberg, 1990). It is possible that most of the more global intonation patterns are composed of pitch range targets similar to those for focus and topic initiation. It awaits future research to explore this possibility.

6. Concluding remarks

To summarize, we have proposed a preliminary framework for accounting for certain surface F_0 variations in speech. The framework consists of definitions for pitch targets and rules of their implementation. Pitch targets are defined as the smallest articulatorially operable units associated with linguistically functional pitch units, and are comparable to segmental phones. They are either static or dynamic. The implementation rules are based on possible articulatory constraints on the production of surface F_0 contours. Due to these constraints, the implementation of a simple pitch target may result in surface F_0 forms that only partially reflect the underlying pitch targets. We have also discussed possible implications of this framework on our understanding of various observed F_0 patterns, including carryover and anticipatory variations, downstep, declination, and F_0 peak alignment. Finally, we considered possible interactions between local and non-local pitch targets.

The framework in its current form is still fairly preliminary. Many details about the specific constraints need to be filled in; pitch targets related to both local pitch contours and global intonation patterns in various languages need to be investigated; and the host units that carry the pitch targets need to be determined for different languages. Furthermore, the framework as presented in this paper is qualitative. To further test its validity, we need to develop a quantitative model that incorporates the principles proposed in the framework. Such was attempted in a recent study (Xu et al., 1999), and the initial results appeared promising.

Acknowledgements

We would like to thank Cinzia Avesani, Ann Bradlow, Chilin Shih, and an anonymous reviewer for valuable comments on earlier versions of our paper.

References

Abramson, A.S. 1978. The phonetic plausibility of the segmentation of tones in Thai phonology. In: Proceedings of the 12th International Congress of Linguistics, Vienna.

Abramson, A.S., 1979. The coarticulation of tones: An acoustic study of Thai. In: Thongkum, T.L., Kullavanijaya, P., Panupong, V., Tingsabadh, K. (Eds.), Studies in Tai and Mon-Khmer Phonetics and Phonology in Honour of Eugenie J.A. Henderson. Chulalongkorn University Press, Bangkok, pp. 1–9.

Anderson, S.R., 1978. Tone features. In: Fromkin, V.A. (Ed.), Tone: A Linguistic Survey. Academic Press, New York, pp. 133–175.

Arvaniti, A., Ladd, D.R. 1995. Tonal alignment and the representation of accentual targets. In: Proceedings of the 13th International Congress – Phonetic Science, Vol. 4, Stockholm, pp. 220–223.

Arvaniti, A., Ladd, D.R., Mennen, I., 1998. Stability of tonal alignment: the case of Greek prenuclear accents. J. Phonetics 36, 3–25.

Bolinger, D., 1989. Intonation and Its Uses – Melody in Grammar and Discourse. Stanford University Press, Stanford, CA.

Bruce, G., 1977. Swedish Word Accents in Sentence Perspective. Gleerup, Lund.

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

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

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

Clark, M. 1978. A dynamic treatment of tone, with special attention to the tonal system of Igbo. Ph.D. dissertation, University of Massachusetts.

Clements, G.N., Ford, K., 1979. Kikuyu tone shift and its synchronic consequences. Linguistic Inquiry 10, 179–210.

Cohen, A., Collier, R., 't Hart, J., 1982. Declination: Construct or intrinsic feature of speech pitch. Phonetica 39, 254–273.

Collier, R., 1987. F_0 declination: The control of its setting, resetting, and slope. In: Baer, T., Sasaki, C.T., Harris, K.S. (Eds.), Laryngeal Function in Phonation and Respiration. College-Hill Press, Boston, pp. 403–421.

Cooper, W.E., Eady, S.J., Mueller, P.R., 1985. Acoustical aspects of contrastive stress in question–answer contexts. J. Acoust. Soc. Amer. 77, 2142–2156.

Crystal, D., 1969. Prosodic Systems and Intonation in English. Cambridge University Press, London.

de Jong, K., 1994. Initial tones and prominence in Seoul Korean. OSU Working Papers in Linguistics 43, 1–14.

Duanmu, S., 1994. Against contour tone units. Linguistic Inquiry 25, 555–608.

Eady, S.J., Cooper, W.E., 1986. Speech intonation and focus location in matched statements and questions. J. Acoust. Soc. Amer. 80, 402–416.

Eady, S.J., Cooper, W.E., Klouda, G.V., Mueller, P.R., Lotts, D.W., 1986. Acoustic characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. Language Speech 29, 233–251.

Fujisaki, H., 1988. A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In: Fujimura, O. (Ed.), Vocal Physiology: Voice Production. Raven Press, New York, pp. 347–355.

Gandour, J., 1974. On the representation of tone in Siamese. UCLA Working Papers in Phonetics 27, 118–146.

Gandour, J., Potisuk, S., Dechongkit, S., 1994. Tonal coarticulation in Thai. J. Phonetics 22, 477–492.

Gårding, E., 1979. Sentence intonation in Swedish. Phonetica 36, 207–215.

Gårding, E., 1987. Speech act and tonal pattern in standard Chinese. Phonetica 44, 13–29.

Gelfer, C.E., Harris, K.S., Collier, R., Baer, T., 1985. Is declination actively controlled?. In: Titze, I.R., Scherer, R.C. (Eds.), Vocal Fold Physiology: Biomechanics and Phonatory Control. Denver Center for the Performing Arts, Denver, CO, pp. 113–126.

Goldsmith, J.A., 1979. Autosegmental Phonology. Garland Press, New York.

Goldsmith, J.A., 1990. Autosegmental and Metrical Phonology. Blackwell Publishers, Oxford.

Grimm, C., 1997. Pitch accent in Oneida. Presentation at the 1997 Annual Meeting of the Linguistic Society of America, Chicago.

Grønnum, N. 1995. Superposition and subordination in intonation – a non-linear approach. In: Proceedings of the 13th International Congress – Phonetic Science, Vol. 2, Stockholm, pp. 124–131.

Hirschberg, J., Ward, G., 1992. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise–fall–rise intonation contour in English. J. Phonetics 20, 241–251.

Hombert, J.-M. 1974. Universals of downdrift: their phonetic basis and significance for a theory of tone. Studies in African Linguistics (Suppl. 5) 169–183.

Hombert, J.-M., 1978. Consonant types, vowel, quality and tone. In: Fromkin, V.A. (Ed.), Tone: A Linguistic Survey. Academic Press, New York, pp. 77–111.

Howie, J.M., 1974. On the domain of tone in Mandarin. Phonetica 30, 129–148.

Hyman, L.M., 1993. Register tones and tonal geometry. In: Hulst, H.v.d., Snider, K. (Eds.), The Phonology of Tone. Mouton de Gruyter, New York, pp. 75–108.

Hyman, L., Schuh, R., 1974. Universals of tone rules. Linguistic Inquiry 5, 81–115.

Jin, S. 1996. An acoustic study of sentence stress in Mandarin Chinese. Ph.D. dissertation, Department of East Asian Language and Literature, The Ohio State University.

Kelso, J.A.S., 1984. Phase transitions and critical behavior in human bimanual coordination. Am. J. Physiol.: Regulatory, Intergrative Comparative 246, R1000–R1004.

Ladd, D.R., 1983. Phonological features of intonational peaks. Language 59, 721–759.

Ladd, D.R., 1984. Declination: A review and some hypothesis. Phonology Yearbook 1, 53–74.

Ladd, D.R. 1995. "Linear" and "overlay" descriptions: An autosegmental–metrical middle way. In: Proceedings of the 13th International Congress – Phonetic Science, Vol. 2, Stockholm, pp. 116-123.

Ladd, D.R., 1996. Intonational Phonology. Cambridge University Press, Cambridge.

Ladd, D.R., Silverman, K.E.A., 1984. Vowel intrinsic pitch in connected speech. Phonetica 41, 31–40.

Laniran, Y. 1992. Intonation in tone languages: The phonetic implementation of tones in Yorùbá. Ph.D. dissertation, Cornell University.

Laniran, Y., Gerfen, C. 1997. High raising, downstep and downdrift in Igbo. Presentation at the 71st Annual Meeting of the Linguistic Society of America, Chicago.

Leben, W.R. 1973. Suprasegmental phonology, Ph.D. dissertation, Massachusetts Institute of Technology.

Lehiste, I., 1975. The phonetic structure of paragraphs. In: Cohen, A., Nooteboom, S.E.G. (Eds.), Structure and Process in Speech Perception. Springer, New York, pp. 195–206.

Lehiste, I., Peterson, G.E., 1961. Some basic considerations in the analysis of intonation. J. Acoust. Soc. Amer. 33, 419–425.

Liberman, M., Pierrehumbert, J., 1984. Intonational invariance under changes in pitch range and length. In: Aronoff, M., Oehrle, R. (Eds.), Language Sound Structure. MIT Press, Cambridge, MA, pp. 157–233.

Lieberman, P., 1967. Intonation, Perception and Language. MIT Press, Cambridge, MA.

Lin, M., Yan, J. 1991. Tonal coarticulation patterns in quadrasyllabic words and phrases of Mandarin. In: Proceedings of the 12th International Congress – Phonetic Science Aix-en-Provence, France, pp. 242–245.

Lindblom, B., 1963. Spectrographic study of vowel reduction. J. Acoust. Soc. Amer. 35, 1773–1781.

Maeda, S., 1976. A Characterization of American English Intonation. MIT Press, Cambridge, MA.

Manfredi, V., 1993. Spreading and downstep: Prosodic government in tone languages. In: Hulst, H.v.d., Snider, K. (Eds.), The Phonology of Tone. Mouton de Gruyter, New York, pp. 133–184.

Mattingly, I.G., 1966. Synthesis by rule of prosodic features. Language Speech 9, 1–13.

Nakajima, S., Allen, J.F., 1993. A study on prosody and discourse structure in cooperative dialogues. Phonetica 50, 197–210.

O'Connor, J.D., Arnold, G.F., 1961. Intonation of Colloquial English. Longmans, London.

Ohala, J.J. 1990. Respiratory activity in speech. In: Hardcastle, Marchal (Eds.), Speech Production and Speech Modelling. Kluwer Academic Publishers, Dordrecht, pp. 23–53.

Ohala, J.J., Ewan, W.G., 1973. Speed of pitch change. J. Acoust. Soc. Amer. 53, 345A.

Öhman, S.E.G, 1966. Coarticulation in VCV utterances: Spectrographic measurements. J. Acoust. Soc. Amer. 39, 151–168.

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

Pierrehumbert, J., Hirschberg, J., 1990. The meaning of intonational contours in the interpretation of discourse. In: Cohen, P.R., Morgan, J., Pollack, M.E. (Eds.), Intentions in Communication. MIT Press, Cambridge, MA, pp. 271–311.

Pierrehumbert, J., Steele, S.A., 1989. Categories of tonal alignment in English. Phonetica 46, 181–196.

Pike, K.L., 1948. Tone Languages. University of Michigan Press, Ann Arbor.

Poser, W. 1984. The phonetics and phonology of tone and intonation in Japanese. Ph.D. dissertation, Massachusetts Institute of Technology.

Prieto, P., van Santen, J., Hirshberg, J., 1995. Tonal alignment patterns in Spanish. J. Phonetics 23, 429–451.

Prieto, P., Shih, C.-L., Nibert, H., 1996. Pitch downtrend in Spanish. J. Phonetics 24, 445–473.

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

Schmidt, R.C., Carello, C., Turvey, M.T., 1990. Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. J. Exp. Psychol.: Human Perception Performance 16, 227–247.

Schuh, R.G., 1978. Tone rules. In: Fromkin, V.A. (Ed.), Tone: A Linguistic Survey. Academic Press, New York, pp. 221–256.

Shen, X.S., 1990. Tonal coarticulation in Mandarin. J. Phonetics 18, 281–295.

Shi, B., Zhang, J. 1987. Vowel intrinsic pitch in Standard Chinese. In: Proceedings of the 11th International Congress – Phonetic Science, Tallinn, Estonia, pp. 142–145.

Shih, C.-L., 1988. Tone and intonation in Mandarin. Working Papers, Cornell Phonetics Laboratory 3, 83–109.

Shih, C.-L., 1997. Declination in Mandarin. In: Botinis, A., Kouroupetroglou, G., Carayannis, G. (Eds.), Intonation: Theory, Models and Applications, Proceedings of an ESCA Workshop. European Speech Communication Association, Athens, Greece, pp. 293–296.

Silverman, K.E.A., Pierrehumbert, J.B., 1990. The timing of prenuclear high accents in English. In: Kingston, J., Beckman, M.E. (Eds.), Papers in Laboratory Phonology 1 – Between the Grammar and Physics of Speech. Cambridge University Press, Cambridge, pp. 72–106.

Steele, S.A., 1986a. Interaction of vowel F_0 and prosody. Phonetica 43, 92–105.

Steele, S.A., 1986b. Nuclear accent F_0 peak location: Effects of rate, vowel, and number of following syllables. J. Acoust. Soc. Amer. 80, S51.

Stevens, K.N., House, A.S., 1963. Perturbation of vowel articulations by consonantal context: An acoustical study. J. Speech Hearing Res. 6, 111–128.

Stewart, J.M., 1983. Key lowering (downstep/downglide) in Dschang. J. African Languages Linguistics 3, 113–138.

Sundberg, J., 1979. Maximum speed of pitch changes in singers and untrained subjects. J. Phonetics 7, 71–79.

Titze, I.R., Durham, P.L., 1987. Passive mechanisms influencing fundamental frequency control. In: Baer, T., Sasaki, C.T., Harris, K.S. (Eds.), Laryngeal Function in Phonation and Respiration. College-Hill Press, Boston, pp. 304–319.

Umeda, N., 1982. F_0 declination is situation dependent. J. Phonetics 10, 279–290.

van Santen, J.P.H., Hirschberg, J., 1994. Segmental effects on timing and height of pitch contours. In: Proceedings of 1994 International Conference on Spoken Language Processing. Yokohama, Japan, pp. 719–722.

Wang, W.S.Y., 1967. Phonological features of tone. Int. J. Am. Linguistics 33, 93–105.

Whalen, D.H., Levitt, A.G., 1995. The universality of intrinsic F_0 of vowels. J. Phonetics 23, 349–366.

Woo, N. 1969. Prosody and phonology. Ph.D. dissertation, Massachusetts Institute of Technology.

Xu, C.X., Xu, Y., Luo, L.-S. 1999. A pitch target approximation model for F_0 contours in Mandarin. In: Proceedings of the 14th International Congress – Phonetic Science, San Francisco, pp. 2359–2362.

Xu, Y. 1993. Contextual tonal variation in Mandarin Chinese. Ph.D. dissertation, The University of Connecticut.

Xu, Y., 1994. Production and perception of coarticulated tones. J. Acoust. Soc. Amer. 95, 2240–2253.

Xu, Y., 1997. Contextual tonal variations in Mandarin. J. 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., 1999a. Effects of tone and focus on the formation and alignment of F_0 contours. J. Phonetics 27, 55–105.

Xu, Y. 1999b. F_0 peak delay: When, where and why it occurs. In: Proceedings of the 14th International Congress – Phonetic Science, San Francisco, pp. 1881–1884.

Xu, Y., Wang, Q.E. 1997. What can tone studies tell us about intonation? In: Botinis, A., Kouroupetroglou, G., Carayannis, G. (Eds.), Proceedings of an ESCA Workshop on Intonation: Theory, Models and Applications, Athens, pp. 337–340.

Yip, M., 1991. The Tonal Phonology of Chinese. Garland Publishing, New York.