

## Production of Weak Elements in Speech – Evidence from F<sub>0</sub> Patterns of Neutral Tone in Standard Chinese

Yiya Chen<sup>a</sup> Yi Xu<sup>b,c</sup>

<sup>a</sup>Radboud University Nijmegen, The Netherlands; <sup>b</sup>University College London, London, UK; <sup>c</sup>Haskins Laboratories, New Haven, Conn., USA

### Abstract

Many weak elements in speech, such as schwa in English and neutral tone in Standard Chinese, are commonly assumed to be unspecified or underspecified phonologically. The surface phonetic values of these elements are assumed to derive from interpolation between the adjacent phonologically specified elements or from the spreading of the contextual phonological features. In the present study, we re-evaluate this view by investigating detailed F<sub>0</sub> contours of neutral-tone syllables in Standard Chinese, which are widely accepted as toneless underlyingly. We recorded sentences containing 0–3 consecutive neutral-tone syllables at two speaking rates with two focus conditions. Results of the experiment indicate that neutral-tone syllables do have a target that is independent of the surrounding tones, which is likely to be static and mid. Furthermore, the neutral tone is found to be different from the full lexical tones in the manner with which the underlying tonal target is implemented: it is slow and ineffective both in overcoming the influence of the preceding full lexical tone and in approaching its own target. Applying the recently proposed pitch target approximation model, we conclude that the neutral tone differs from the other lexical tones in Standard Chinese not only in terms of its mid target, but also in terms of the weak articulatory strength with which this target is implemented. Finally, we suggest that this new understanding is potentially applicable to other weak elements in speech.

Copyright © 2006 S. Karger AG, Basel

### 1 Introduction

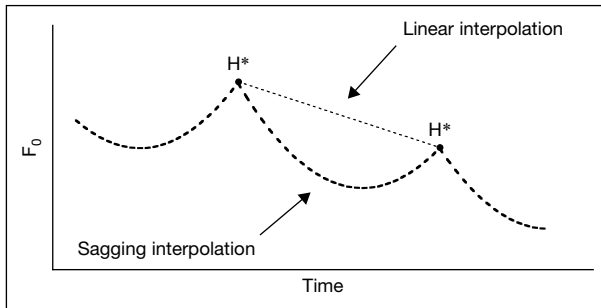
A commonly assumed view of phonology/phonetic interface is that at the end of phonological derivation, phonological features, abstract and discrete, are converted to phonetic *targets*, spatial and temporal, and then *connected* in the physical realizations of speech [Keating, 1988a]. The mechanisms employed to convert and connect the targets are called phonetic implementation rules. Under such a model, the tasks of phonetics are two-fold: *target realization* and *target connection*. This view of phonetic

#### KARGER

Fax +41 61 306 12 34  
E-Mail [karger@karger.ch](mailto:karger@karger.ch)  
[www.karger.com](http://www.karger.com)

© 2006 S. Karger AG, Basel  
0031–8388/06/0631–0047  
\$23.50/0  
Accessible online at:  
[www.karger.com/journals/pho](http://www.karger.com/journals/pho)

Yiya Chen  
Radboud University  
PO Box 9103, NL–6500 HD Nijmegen (The Netherlands)  
Tel. +31 24 3611075, Fax +31 24 3611070  
E-Mail [yiya.chen@let.ru.nl](mailto:yiya.chen@let.ru.nl)



**Fig. 1.** Illustration of ‘target connection’ through linear or ‘sagging’ interpolation between two H\* accents as proposed in the AM theory of English intonation [Pierrehumbert, 1980; Beckman and Pierrehumbert, 1986]. The ‘sagging’ interpolation is drawn as a parabolic function:  $F_0 = at^2 + bt + c$ , used in Pierrehumbert [1981]. The graph is adapted from Xu and Xu [2005].

implementation of phonological features is made most explicit regarding the generation of  $F_0$  patterns related to tone and intonation. Targets are assumed to be realized as turning points in  $F_0$  contours, while target connection provides surface  $F_0$  values between the targets. As is schematically illustrated in figure 1, two types of connections have been proposed: linear connection and sagging interpolation [Bruce, 1977; Pierrehumbert, 1980; Pierrehumbert and Beckman, 1988].

Given such an understanding, it is natural to hypothesize that only the part of the acoustic signals directly reflecting the phonological features is fully specified, although special phonological processes are sometimes posited to map the hypothesized phonological features onto the varied phonetic targets (e.g. the tone sandhi rules in the case of lexical tones in Chen [2000] and references therein, or the phonetic implementation rules in the case of English intonation cited above). The rest of the acoustic signal is then presumed to be unspecified for phonologically distinctive features (at the segmental level, see Cohn [1993] and Huffman [1993] on nasal, Choi [1995] on vowels, and Keating [1988b] on consonant place of articulation; at the suprasegmental level, see Pierrehumbert [1980], Pierrehumbert and Beckman [1988] and most of their subsequent work on intonation, Shih [1987] on tone in Mandarin, and Myers [1999] on tone in Chichewa). Furthermore, regions unspecified for phonological features are filled with phonetic values through *interpolation* between adjacent targets, as illustrated by the dashed lines in figure 1.

Such a model of phonetics/phonology interface has met some challenges in recent years. One particular kind of challenge comes from the finding that certain segments not believed to be specified for any feature are nevertheless realized with values that cannot be readily explained by interpolation. Rather, they seem to exhibit characteristics similar to feature-specified targets. Browman and Goldstein [1992], for example, show that contextual vowel qualities are not sufficient to predict the surface vowel quality of the schwa. They illustrate with modelling data that schwa must also have an underlying target. This view is supported by Gick [2002], who shows that the production of schwa, as demonstrated by X-ray data, indeed must have an underlying gestural target. Similarly, the epenthetic schwa in Dutch has also been argued to have a phonologically specified unit [Warner et al., 2001]. Boyce et al. [1992], with articulatory movement data for the lips and the velum, show that when given enough time, segments which are not specified for certain features nevertheless exhibit articulatory movements characteristic of those

features (such as lip rounding for an alveolar [t]). These findings motivate the rethinking of the dividing line between phonological features and phonetic targets, or the conversion of phonological features in phonetic targets.

There are also challenges in studies of intonation. As mentioned earlier, the distinctive tonal features are realized as phonetic targets in the form of  $F_0$  turning points, and that the rest of the  $F_0$  contour comes from interpolation between the turning points [Bruce, 1977; Pierrehumbert, 1980, 1981].<sup>1</sup> The unmarked pattern of interpolation is assumed to be linear, but two additional mechanisms have been proposed to account for cases of apparent nonlinear connection between  $F_0$  turning points when tonal targets are presumably sparsely distributed in time: target spreading and sagging interpolation [Pierrehumbert, 1980, 1981]. Of particular interest here is the sagging interpolation, proposed first to account for the dipping contour between two  $H^*$  tones in English. Such an interpolation is claimed to differ from that between an  $H^*$  and a following  $L+H^*$  tone in English intonation, which is assumed to be linear. Furthermore, the valley observed between two  $H^*$  tones is claimed not to have a phonological status as that of the  $L$  in an  $L+H^*$  tone sequence [Pierrehumbert, 1980]. Myers [1999] applies this view to a tone language, arguing that in Chichewa, a language with two level tones ( $H$  and  $L$ ), the  $L$  tone is phonologically underspecified, and that the surface  $F_0$  pattern of the  $L$  tone comes from sagging transitions between adjacent  $H$  tones.

The problem with the proposal of ‘sagging interpolation’ is that it admits the existence of a covert factor that is partially independent of the specification of the flanking targets, because otherwise there is no reason for an interpolation line to ‘sag’. The nature of this independent factor thus calls for an explanation. Ladd and Schepman [2003], based on evidence from both production and perception experiments on English intonation, argue that the  $F_0$  minimum between two high accents (in the cases that they have examined) in English ‘corresponds to a phonologically specified  $L$  tone’ (p. 104), which would eliminate the need for a sagging transition between  $H^*$  pitch accents in English.<sup>2</sup> Xu and Xu [2005] argue that the  $F_0$  valley described as  $L$  by Ladd and Schepman should be associated with the unaccented syllables preceding the accented syllable, because its height is not correlated with the  $F_0$  peak of the upcoming accent. The status of  $F_0$  valleys before an  $H^*$  pitch accent is also questioned by Auferbeck [2002] which examines the  $F_0$  pattern from the beginning of an intonational phrase to the first  $H^*$  pitch accent of the phrase. It is found that  $F_0$  stays flat on a low pitch level until the onset of the  $H^*$  pitch accent, despite the increasing number of unaccented syllables before the onset of the  $H^*$ .

In light of these new yet somewhat inconclusive findings on the nature of targets as well as the mechanisms for target realization and connection, it has become clear that new data, preferably from different languages, are needed to further clarify these issues. In this paper, we present results of a detailed phonetic study on the  $F_0$  realization of the neutral tone in Standard Chinese, which is conventionally believed to be toneless, i.e. targetless, just as the schwa in English. Our goal is not only to improve our understanding of the relation between the underlying and surface forms of the neutral

<sup>1</sup> Models like the British School of Intonation [Cruttenden, 1997; Crystal, 1969; O’Connor and Arnold, 1961] and the Kiel Intonation Model [Kohler, 1997, 2004] assume that the essential components of intonation are the entire contours such as  $F_0$  peaks and valleys rather than pitch registers in the form of  $F_0$  turning points. These models therefore assume that the  $F_0$  trajectories between the turning points are also specified underlyingly.

<sup>2</sup> See also the recognition of a low tonal target in describing similar  $F_0$  patterns for English [Gussenhoven, 1983], for Italian [D’Imperio, 1999], and for Spanish [Face, 2001].

tone in Standard Chinese, but also to bring new insight into the relation between underlying and surface forms of ‘weak’ elements in speech in general.

## 2 The Neutral Tone in Standard Chinese

In Standard Chinese, besides the four lexically distinctive full tones (H = high level, R = rising, L = falling-rising, F = falling), there exist a number of items described by the cover term neutral tone [Chao, 1968]. Typical examples include grammatical morphemes (e.g. the genitive/nominalizer marker *de* in 1a), lexical items (*li* in 1b), diminutive terms (*mei* in 1c), and reduplication (the second *xiang* in 1d). (Syllables without tonal marking indicate neutral tone.)

- (1) (a) *làde*            ‘something spicy’  
      (b) *bōli*            ‘glass’  
      (c) *mèimei*        ‘sister (diminutive)’  
      (d) *xiangxiang*    ‘to think (for a little while)’

What these syllables have in common is that they do not occur in initial positions and their surface  $F_0$  contours are much less consistent than those of the full-tone syllables. In fact, their  $F_0$  realization has been believed to be completely dependent on the tone of the preceding syllable, as described impressionistically in the literature [e.g. Chao, 1968; Cheng, 1973]. Early acoustic studies also confirmed that the  $F_0$  contour of the neutral tone changes according to the tone of the preceding syllable [e.g. Lin and Yan, 1980]. The contextual dependency of the neutral-tone  $F_0$  contour and the fact that most neutral-tone morphemes have full-tone alternates have led to the general consensus that there are no phonological specifications for the neutral tone (however, see Yip [1980] who posits that the neutral tone is specified for [-upper] register and Lin [2001] who posits a low tone for the neutral tone).

Different proposals have been put forward as to how surface  $F_0$  contours of the neutral tone are realized. Yip [1980] proposed that the  $F_0$  contour of the neutral tone is realized by tonal spreading from the preceding syllable. Shen [1992] adopted this view and proposed further that the ending component of the preceding tone was the neutral tone target. Shih [1987] proposed that the neutral tone in sentence medial position, in certain tonal contexts, is better accounted for by interpolation. For the neutral tone in other contexts, she posited extra tonal targets in order to explain the observed contours. van Santen et al. [1998] also proposed that the neutral tone has a mid target placed at a location somewhere before the end of the syllable,<sup>3</sup> and its  $F_0$  comes from interpolation between the preceding full tone and this mid target.<sup>4</sup> Wang [1997] accounted for the surface  $F_0$  realization of the neutral tone by stipulating a complicated set of phonological and phonetic rules. None of these studies, however, described the  $F_0$  contours of the neutral tones in detail. Rather, the conclusions were based either on informal observation or on a few measurement points of the  $F_0$  contours. Li [2003] obtains experimental

<sup>3</sup> They also proposed that the target of the neutral tone is changed from M to L when following a falling tone.

<sup>4</sup> Although the discussion there did not specify that the surface  $F_0$  of the neutral tone comes from interpolation, earlier in the same paper it was explained that the synthesis of Mandarin was done by applying ‘a Pierrehumbert-style tone sequence approach’ [van Santen et al., 1998, p. 151].

data in much more detail than previous studies. To account for the observed surface  $F_0$  patterns, he posited that the neutral-tone syllables form a prosodic word with the preceding full-tone syllable. Furthermore, for each prosodic word, there is a boundary tone. The surface  $F_0$  over the neutral-tone syllable(s) is then derived via interpolation between the preceding lexical tone and the boundary tone. An underlying assumption is that the neutral-tone syllables always form a prosodic word with their preceding full lexical tone syllable(s) irrespective of the varied morphosyntactic structures they may form.

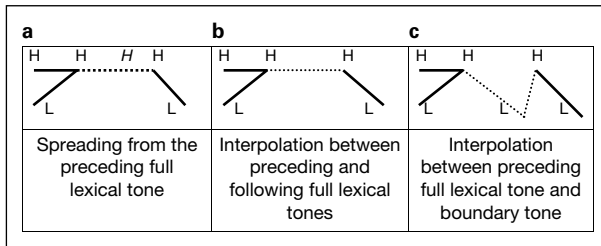
The assumption shared by the above-mentioned accounts is that the surface  $F_0$  of the neutral tone comes from target interpolation. That is, the adjacent full tones are first realized as  $F_0$  targets, and the neutral tone either acquires (or further specifies) a target from the spreading of the preceding tone [Yip, 1980; Wang, 1997], or its  $F_0$  value is obtained simply via interpolation between the surface targets of the preceding full lexical tone and that of the following tone, which could be the following full lexical tone [Shih, 1987] or a postlexical prosodic word boundary tone [Li, 2003].

Such an assumption, however, goes against the general findings about  $F_0$  contour formation in full lexical tones. For interpolation to work, whether it is linear or not, both a starting point and an ending point have to be prespecified. But it has been established in recent research that in connected speech, the onset  $F_0$  of any tone varies extensively with the offset  $F_0$  of the preceding tone [Gandour et al., 1994, for Thai; Xu, 1997, 1999, for Standard Chinese; Li and Lee, 2002, for Cantonese]. If the  $F_0$  onset of an upcoming full tone is not fixed, interpolation is virtually impossible.

Another possibility is suggested by Xu and his colleagues [Xu 1997, 1999; Xu and Wang, 2001]. They argue that the process of producing a tone is to asymptotically approach its underlying target within the time interval allocated to it, which is usually the duration of the tone-carrying syllable. An underlying target is defined as a simple linear function that is either static (i.e., slope = 0) or dynamic (i.e., slope  $\neq$  0). The tonal implementation is understood as being done under various articulatory constraints, among which the most critical are the maximum speed of pitch change [Xu and Sun, 2002] and the synchronization of laryngeal and supralaryngeal movements [Xu and Wang, 2001]. Under these articulatory constraints, the underlying pitch targets are often not fully realized. Nonetheless, the consistency of each tone is manifested in the continuous convergence to its underlying pitch target over the time interval allocated to the tone (i.e. the syllable duration). This new model of tone production thus assumes *target approximation* rather than target interpolation as the basic mechanism of realizing consecutive lexical tones. In such a model, *targets* are the *intended* goals rather than the realized  $F_0$  points or contours, and a target is required for each and every tone, including the neutral tone, for there is no assumed mechanism with which the  $F_0$  of *any* tone can be fully generated by other tones.

This target approximation model predicts that the greater surface variability of the neutral tone than that of the full lexical tones could result simply from the fact that the neutral-tone syllables are usually short – almost half as long as the full-tone syllables [Lin and Yan, 1980]. That is, other things being equal, less time would lead to less adequate target attainment. On the other hand, if time does become more abundant, the underlying target for the neutral tone, if there is one, should become more apparent.

Thus, the applicability of the target approximation model to the neutral tone is testable if we adopt a method similar to that employed by Boyce et al. [1992] in a study on lip protrusion and velum movement. They showed that by increasing the number of



**Fig. 2.** Predictions of  $F_0$  contours of neutral tone(s) between adjacent full tones (HH or LH as the preceding tone and HL as the following tone) by the tone spreading and linear interpolation hypotheses.

segments that are ‘unspecified’ for the targeted phonological features or by decreasing speaking rate, independent articulatory targets for these features can be observed. Their results suggest that the variability in timing indeed helps in ‘bringing out’ articulatory targets that are characteristic of noncontrastive features (i.e. rounding and nasal).

A further benefit of using this method is that, if interpolation is a viable mechanism, the extra time provided by the increased number of neutral-tone syllables should allow the path of  $F_0$  interpolation to be manifested more clearly, as is shown in figure 2, which depicts schematically the predictions existing proposals would make on the  $F_0$  contours of the neutral-tone syllables preceded and followed by syllables with different lexical full tones. In the figure, the full tones are labeled as L and H, as is conventionally done in the prior studies.

Figure 2a gives the schematic contours predicted by tonal spreading (from the preceding LH or HH tone). Here the increase in the number of neutral-tone syllables results in an increased number of high tones (as indicated by the italic H) and therefore a longer stretch of high-level  $F_0$ . The transition toward the falling tone (HL) starts at the end of the last neutral-tone syllable. Figure 2b gives the schematic contours predicted by interpolation between adjacent full tones. Here the increase in the number of neutral-tone syllables results in a longer flat line between the two  $F_0$  heights since unspecified elements, namely the targetless neutral tones, should not contribute anything of their own to the trajectory of the  $F_0$  contour. Figure 2c illustrates the schematic contours predicted by interpolation between the preceding full tone and an intonational low boundary tone inserted at the end of the last neutral-tone syllable. Here the predicted contour would drop over the course of the neutral-tone syllable(s), but the end point of the drop would remain constant irrespective of the number of neutral syllables in the sequence.

The prediction of the target approximation model for the neutral tone will be different from the above predictions. First, although the target of the neutral tone is not yet known, it is unlikely to be as high as that of the H tone or as low as that of the L tone, since those would be close to the extremes of the tonal pitch range. Thus, it is unlikely that the  $F_0$  of the neutral tone would stay flat between the two high tonal targets. Rather,  $F_0$  is more likely to go down during the neutral-tone syllable and would not go up until after the syllable offset. Second, since the neutral-tone target is approximated, the offset  $F_0$  of the neutral tone would not be fixed. Rather, it would vary with the preceding full tone and the number of consecutive neutral-tone syllables.

We thus designed an experiment to test the predictions of the above proposals in the hope to answer the following question: is the  $F_0$  of the neutral tone best accounted

for by tone spreading, interpolation (between adjacent full tones or toward a boundary tone), or by articulatory approximation (of an underlying target assigned to the neutral tone itself)? By answering this question, we hope to not only reach a better understanding about the neutral tone in Standard Chinese, but also help shed light on the mechanism that underlies the production of weak elements in speech in general.

### 3 Method

#### 3.1 Material

The design of the test material was guided by three observations reported in the literature. First, tonal context, especially the preceding tone, has much influence on the  $F_0$  contour of the neutral tone. So, we varied the tone of the syllable that precedes the neutral-tone syllable. All four lexical tones in Standard Chinese were included (H, which is high level, also known as tone 1; R, which is rising, also known as tone 2; L, which is low in medial positions and low-rising in final and prepausal positions, also known as tone 3; and F, which is falling, also known as tone 4). We also varied the tone of the syllable that follows the neutral-tone syllable(s) to see if there is any contextual effect from the following tone. Here only two lexical tones were included: F, a falling tone that starts high, and L, a low tone that starts low. Second, as mentioned earlier, based on the finding which duration increase can potentially help reveal articulatory targets of noncontrastive phonological features that are otherwise not easily observed [Boyce et al., 1992], we varied both the number of consecutive neutral-tone syllables and speaking rate. Third, because focus introduces considerable changes into the  $F_0$  contours [Jin, 1996; Xu, 1999; Chen, 2003], we also controlled for focus. The test materials thus consist of sentences which vary in (1) the number of consecutive neutral-tone syllables (0–3); (2) tone of the syllable preceding the neutral-tone syllables (4 lexical tones), and (3) tone of the syllable following the last neutral-tone syllable (2 lexical tones). An example is given in 2 (see Appendix A for a full list of the materials).

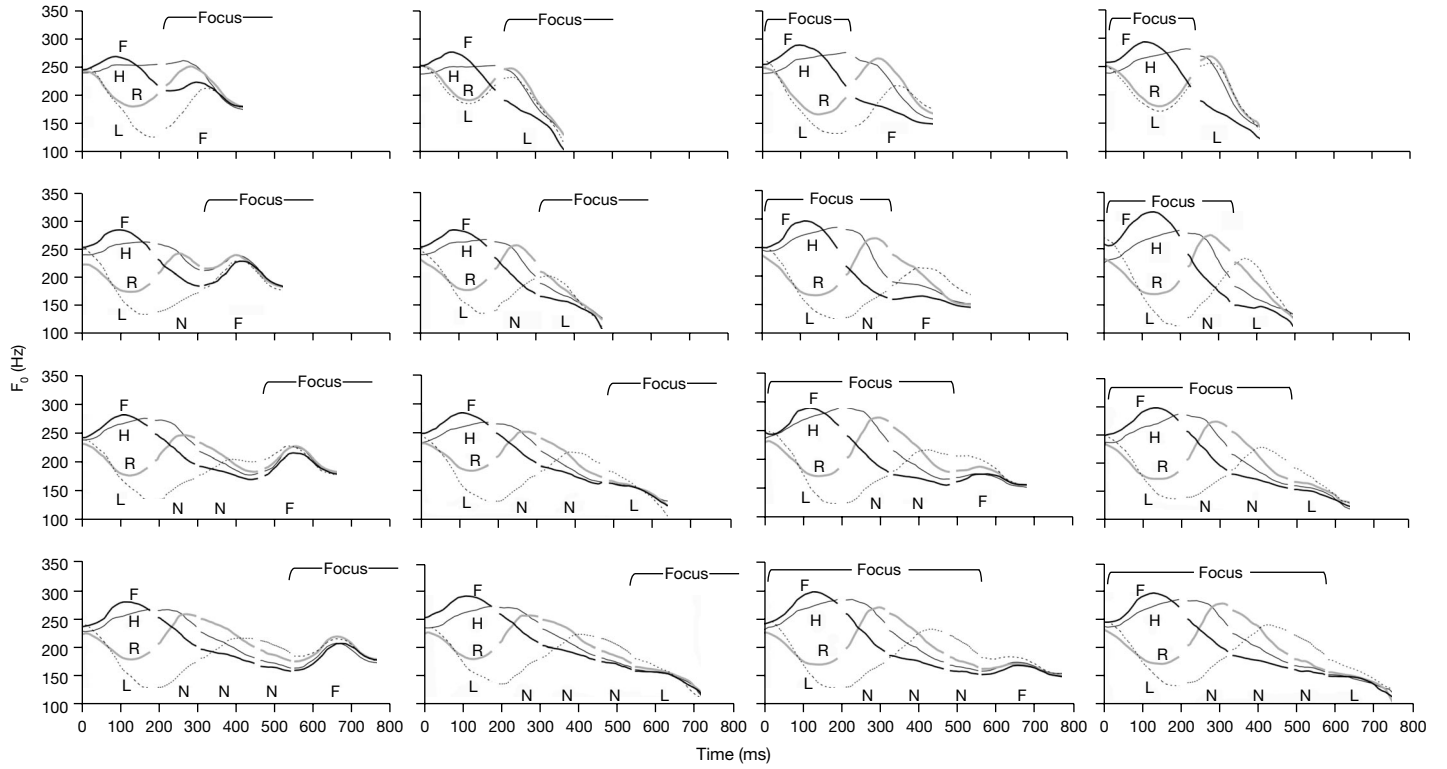
(2)	<u>X</u> <u>Y<sub>0-3</sub></u> <u>Z</u>
	<i>Ta shuo ma ma men de mei duo le</i>
Tone:	H H H N N N L H N
Gloss:	He said moms' beautiful more ASP

Here *X* represents the syllable that precedes the first neutral-tone syllable, *Y* represents the neutral-tone syllables (which vary in number from 1 to 3), and *Z* represents the syllable that follows the last neutral-tone syllable (and in cases where there is no neutral tone, it immediately follows *X*, the first lexical tone syllable). In the subsequent discussion, we will refer to *X* as the first full-tone syllable, *Y* as the neutral-tone syllable, and *Z* as the second full-tone syllable. Furthermore, *X* (and *Y*) forms a noun phrase (or prosodic word according to Li [2003]) serving as the subject of the clause. Two speaking rates were used: normal and fast. To control for the effect of focus, we elicited the utterances as answers to specific questions, which resulted in two focus conditions. In the on-focus condition, focus was on the target noun phrase. In the prefocus condition, focus was on the phrase that follows the target noun phrase, i.e. the adjective plus the comparative adverb in the clause. Thus, there are a total of 32 sentences and 128 different renditions of these sentences:

$$4 (X) \times 4 (Y) \times 2 (Z) \times 2 (\text{focus condition}) \times 2 (\text{speaking rate}) = 128 \text{ renditions}$$

#### 3.2 Subjects

Four speakers of Standard Chinese participated in the experiment (two males and two females). One is the second author and the other three were naïve subjects who were paid for their participation. The paid subjects were all graduate students who were born and raised in Beijing. They were studying at US or UK universities at the time of the recording and had been abroad for different lengths of time but none for more than 5 years. Our second author was not born in Beijing and had been in the US for more than 10 years. But he is a native speaker of Standard Chinese. Data obtained from this study and



previous ones [e.g. Xu, 1999] indicate no apparent phonetic difference between his speech and the rest of the subjects.

### 3.3 Recording

Recordings were conducted in sound-treated booths. One subject was recorded in the Department of Linguistics at Cornell University, two in the Department of Communication Sciences and Disorders, Northwestern University, and one in the Theoretical and Applied Linguistics Department at the University of Edinburgh. Their ages ranged from 25 to 45. All subjects were given the same written instructions and followed the same procedures. All subjects repeated the 128 renditions three times.

### 3.4 $F_0$ Extractions

The  $F_0$  extraction was done using a procedure similar to those used in Xu [1997, 1999]. The procedure combines custom computer programming with ESPS/waves+ (Entropic Inc.). The ESPS *epochs* program was used to mark every vocal period, and the labels were saved into a text file. After that, the waveform, the period labels, and the spectrogram of the signal were displayed in *xwaves*. The period labels were examined for spurious markings such as double tagging and vocal period skipping. Apparent errors were corrected manually. Segment boundaries were hand-labeled. The vocal period and segment labels were then processed by a set of custom-written C programs. The programs converted the vocal periods into  $F_0$  values, and then smoothed the resulting  $F_0$  curves using a trimming algorithm that eliminated sharp bumps and edges [Xu, 1999].

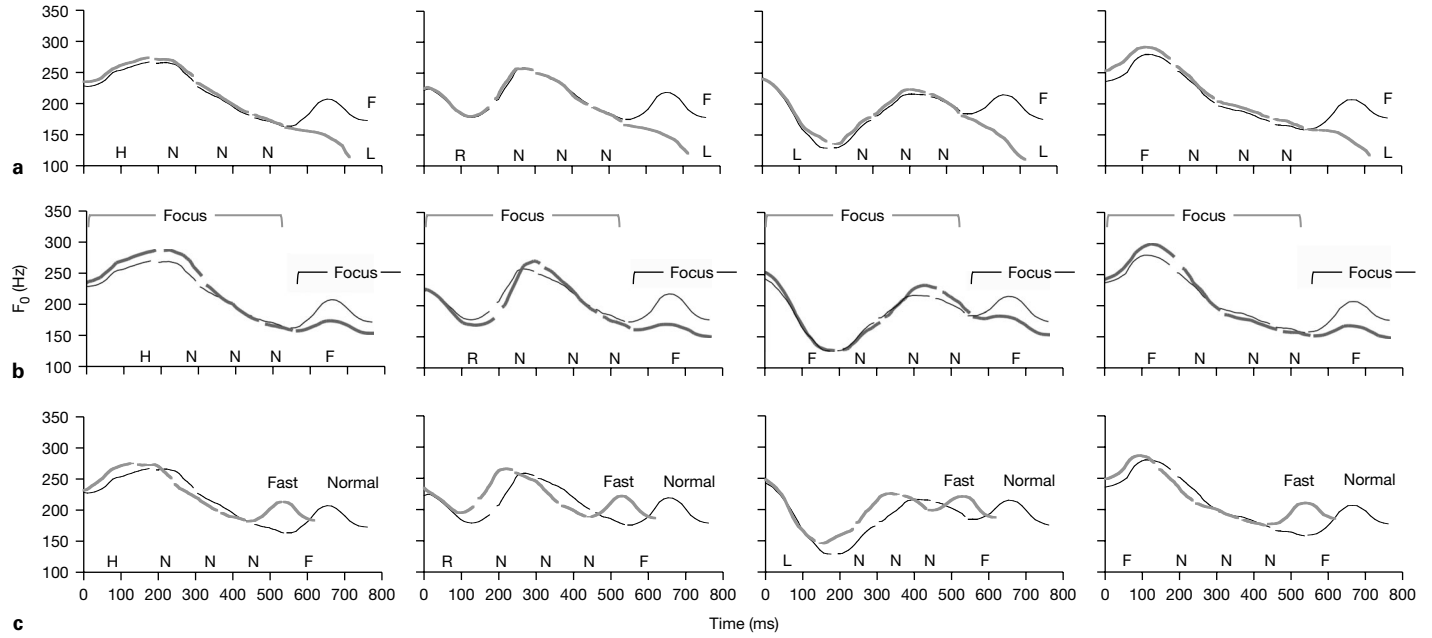
## 4 Analysis

### 4.1 Graphical Comparison of $F_0$ Contours

Figure 3 displays mean  $F_0$  contours of the neutral tone and the surrounding full tones under different focus conditions spoken at normal rate. Each graph represents the  $F_0$  contours of a sequence of syllables: the first full-tone syllable (all four tones), varying number of neutral-tone syllables (0–3), and the second full-tone syllable (falling or low). The  $F_0$  contours in the leftmost two columns (C) were produced in the prefocus condition, i.e. with focus on the phrase that follows the target subject noun phrase. The  $F_0$  contours in the rightmost two columns were produced in the on-focus condition, i.e. with focus on the target subject noun phrase. From top down, the number of neutral-tone syllables increases in each row (R) from 0 to 3.

The  $F_0$  contours in both figures 3 and 4 were obtained by (1) taking 20  $F_0$  points (in hertz) at even intervals from each syllable and (2) averaging them across three repetitions of the same sentence spoken at the same rate and with the same focus pattern.

**Fig. 3.** Mean  $F_0$  contours of neutral-tone syllables in different tonal contexts under different focus conditions. Each curve in a graph is an average of 12 repetitions by 4 subjects, and each gap on a curve indicates a syllable boundary. The four curves in each graph differ in the first full tone, as indicated by H, R, L and F in the first row (R1). The four rows differ in the number of neutral-tone (N) syllables: 0–3 in R1–R4. The second full tone, as indicated by the last tone label in each graph, alternates between F and L across the columns. In the left two columns (C1–C2), the sentence focus is on the phrase starting with the second full-tone syllable, as indicated by ‘focus’ and the half bracket. In the right two columns (C3–C4), the focus is on the word consisting of the first full tone and 0–3 neutral tones, as indicated by ‘focus’ and half bracket around it.



The values were then transformed into semitones before being averaged across speakers and the averaged values were then converted back into hertz for plotting. For ease of visual inspection, the  $F_0$  curves were also normalized in time scale in the following way. First, an average duration was computed of the syllables in each position in sentences having the same number of neutral-tone syllables. Then this averaged duration is used as the time axis for displaying the  $F_0$  contours of each syllable position in the sentence. This way, we can directly compare the tonal contours of different sentences without losing sight of the duration of each syllable.

First, we compared the predicted  $F_0$  contours in figure 2 to the  $F_0$  contours in comparable tonal contexts in figure 3 (C1R4 when the neutral-tone sequence is not focused, C3R4 when the neutral-tone sequence is focused; here their faster counterparts are not shown in the figure). When an H (HH) or an R (LH) tone syllable is followed by a string of 3 neutral-tone syllables and then followed by an F (HL) tone, there is a clear declining pattern of the  $F_0$  contour over the neutral-tone syllable. Such a declining pattern starts during the first neutral-tone syllable and continues as the number of neutral-tone syllables increases (see C1R2-C1R3 and C3R2-C3R3). This is particularly clear when we compare the  $F_0$  contours of H/R N N N (in C1R4) with those of H/R N (C1R2) and H/R N N (C1R3). When there is only one neutral tone (C1R2), a slight dip (about 15–20Hz on average) can be seen in the later part of the  $F_0$  contour of the neutral-tone syllable, although the exact timing of the start of the dip clearly varies according to whether the preceding tone is H or R. When there are two neutral tones (C1R3), we see again a slight dip over the first neutral-tone syllable, but the declining pattern becomes much clearer in the second neutral-tone syllable, similar to the declining  $F_0$  of the second neutral-tone syllable in a 3-neutral-tone sequence in C1R4. This pattern also holds when the following tone is a low tone (if we compare the  $F_0$  contours of neutral-tone syllables in C2/C4R4 to those in C2/C4R2 and C2/C4R3).

Second, the converging value for the neutral tone seems to be around the mid level of the tonal pitch range when the pitch contours of the neutral tone are compared to the following L tone (in C2R2–4 and C4R2–4; see section 4.2.5 for numerical confirmation). Here, the last of the three neutral-tone syllables reaches a value around 150 Hz, with some variations due to the different preceding tones. Then  $F_0$  goes further down to approach the lowest point of the L tone which is known to be the lowest among the four full lexical tones. Furthermore, whatever the underlying target of the neutral tone may be, it is not realized in the same way as the tonal targets of the full lexical tones. In the first row of figure 3, where there is no neutral tone, the influence of the tones from the preceding first full tone is evident in the first half of the following second full-tone syllable. However, the approximation of the second full lexical tone seems to be largely completed by the end of the syllable (for the F tone, see C1R1 and C3R1; for the L tone, see C2R1 and C4R1). In contrast, the approximation of the neutral tone does not seem to be complete even by the end of the third neutral-tone syllable. In other words, the variations due to the tone of the preceding first full-tone syllable, though gradually decreasing in magnitude as the number of neutral-tone syllables increases, remain clearly visible even in the third neutral-tone syllable.

**Fig. 4.** **a** Effects of the second full tone on the preceding neutral tone and the first full tone. **b** Effects of focus. **c** Effects of speaking rate. In each row from left to right, the first full tone is H, R, L and F, respectively.

Third, the carryover effect on the neutral tone is not simply due to the offset  $F_0$  of the preceding first full tone. Rather, the direction of the  $F_0$  movement at the end of the first full tone also plays a role. Firstly, although the offset  $F_0$  of the F tone is about the same as that of the H tone in certain cases (e.g. C4R2, C2R3, C1R4, C2R4 in figure 3), or sometimes even higher than that of the H tone at the faster rate which is not shown in figure 3,  $F_0$  of the following neutral tone in those cases is lower and dropping much more quickly than after the H tone. Secondly, the offset  $F_0$  of the R tone is much lower than that of the H and F tones in all the graphs in figure 3, but the  $F_0$  of the following neutral tone always continues the rising movement of the R tone, resulting in higher  $F_0$  in the later portion of the first neutral-tone syllable than after both the H and F tones in most cases. These patterns seem to make sense if we consider the fact that the realization of the preceding full tone generates an  $F_0$  height as well as a momentum, and the effects of both on the  $F_0$  of the following neutral tone seem to follow the simple physical principle of inertia. Of course, the exact application of this principle is quite complicated in these cases, since muscle and tissue elasticity, compliance and viscosity are all potentially involved.

In contrast to the effects of the other three tones, that of the L tone on the following neutral tone seems to make much less sense in terms of simple physics. The offset of the L tone is the lowest among all four full tones, and  $F_0$  is falling throughout the L-tone syllable until right before the syllable offset when it seems to have leveled off in most cases. But  $F_0$  following the L tone always rises throughout the first neutral-tone syllable, and the rise continues well into the second neutral-tone syllable before eventually reaching a turning point in the middle or later portion of the second neutral-tone syllable (R2–R4). As a consequence, the  $F_0$  at the end of the last neutral-tone syllable is always the highest if the first full tone is L. The implausibility of explanations in simple physics suggests that an additional factor must have been involved. It could be an arbitrary phonological rule, or it could be a physiological mechanism not yet fully recognized. We will return to this issue in our quantitative analyses.

Figure 4 displays the mean  $F_0$  contours of the three-neutral-tone sentences under the effects of the second full tone (fig. 4a), focus (fig. 4b), and speaking rate (fig. 4c). These  $F_0$  contours were obtained with a similar method as those in figure 3. Similar plots for sentences with 0–2 neutral tones were also examined and they were largely the same as those in figure 4. From figure 4a, we can see that the second full tone has little effect on the  $F_0$  of the preceding neutral tones and that of the first full tone: the two  $F_0$  tracks in each graph almost coincide with each other until after the onset of the second full tone. Looking closer at the small differences due to the second full tone, we noticed that they constitute a dissimilatory effect: the  $F_0$  of the preceding neutral tone and even the first full tone was raised by the 2nd full L tone, although the magnitude of the raising is very small. Such ‘anticipatory raising’ is similar to what was previously found in all full-tone sentences [Xu, 1997, 1999].

Figure 4a also shows that the  $F_0$  reached at the end of the third neutral-tone syllable is about halfway between the highest  $F_0$  of the following F tone and the lowest  $F_0$  of the following L tone. This, together with the fact that the second full tone has virtually no assimilatory influence on the preceding neutral tones, indicates that an  $F_0$  value in the middle of the tonal pitch range may be targeted for the neutral tone.

From figure 4b, we can see that the  $F_0$  of the first full tone exhibits an expanded pitch range when it is on focus as compared to the pefocus condition: the mean maximum  $F_0$  of the H, R and F tones are raised, and the mean minimum  $F_0$  of the R tone is lowered. The  $F_0$  of the neutral tone is not much affected except the early portion of the

first neutral tone after H, R and F, which seems to be related to the increased maximum  $F_0$  of the preceding full tone. The  $F_0$  of the second full tone is also clearly affected by focus. It is much lower in the postfocus condition than when it is on focus.

From figure 4c, we can see that faster utterances are shorter than the slower ones, just as expected. But no other effects are immediately apparent in these plots.

## 4.2 Quantitative Analyses

### 4.2.1 Duration

Table 1 displays syllable duration as functions of speaking rate, focus condition and type of second tone. The first two data rows show the duration of the first syllable of the target noun phrase [either monosyllabic or containing neutral-tone syllable(s) after the first full tone], and the last two data rows show the duration of either the second full-tone syllable which follows immediately after the first full-tone syllable when the target noun phrase is monosyllabic or the first neutral-tone syllable in the target noun phrase.

The first two data rows of table 1 show that the duration of the first full-tone syllable in the target noun phrase is shorter when followed by a neutral-tone syllable than by the second full-tone syllable (when the target noun phrase is monosyllabic). On average, the amount of shortening is 34.5 ms, which means that the duration of the pre-neutral-tone syllable is 85% the length of the pre-full-tone syllable. The difference is nonsignificant, however, according to a 5-factor repeated-measures ANOVA [independent variables: first full tone (H/R/L/F), second full tone (L/F), tone type (full/neutral), focus (on-focus/prefocus), and speaking rate (normal/fast)]. Syllable duration in fast speaking rate was shorter than that in normal rate (51.5 ms on average), but the difference is marginally significant [ $F(1, 3) = 13.93, p = 0.0335$ ]. The effect of focus is also marginally significant [ $F(1, 3) = 13.58, p = 0.0346$ ], and the average difference is 22.9 ms.

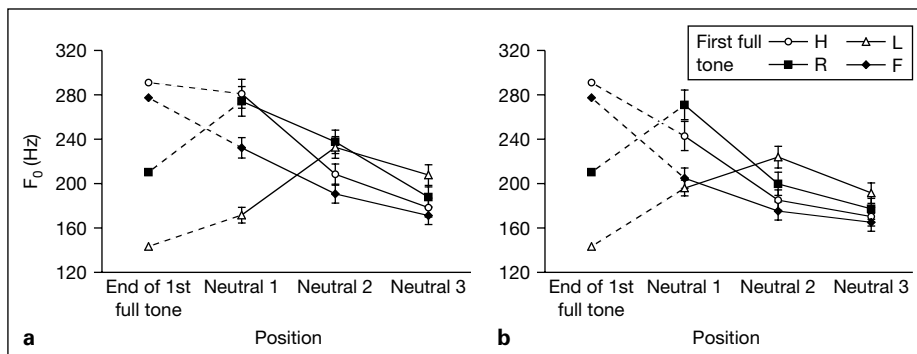
As can be seen in the two bottom rows, the neutral syllables are much shorter than the full tones. On average, the neutral tone is only 61% the length of the full tones (which is slightly longer than what was found by Lin and Yan [1980]). This difference between the neutral tone and the full tones is highly significant according to a 5-factor repeated-measures ANOVA [ $F(1, 3) = 148.27, p = 0.0012$ ] (the same independent variables as in the previous ANOVA). The duration of the second syllable at fast speaking rate is shorter than that at normal rate. But the average difference of 22.1 ms is only marginally significant [ $F(1, 3) = 11.93, p = 0.0408$ ]. The effect of focus is also marginally significant [ $F(1, 3) = 11.72, p = 0.0417$ ], and the average difference is only 3.2 ms.

### 4.2.2 Contextual Influence on Neutral Tone $F_0$

Figure 5 displays mean  $F_0$  values at the middle (fig. 5a) and end (fig. 5b) points of the neutral-tone syllables in sentences with three consecutive neutral tones. The mean  $F_0$  values at the end of the first full tones are also displayed as reference. Here we can see that at the end of the first neutral tone, there is much variance due to the preceding lexical tone; by the end of the second neutral-tone syllable, the  $F_0$  values exhibit a smaller range of variation; by the end of the third neutral-tone syllable, despite some visible influence from the first full tones, the  $F_0$  values have become very close to each other.

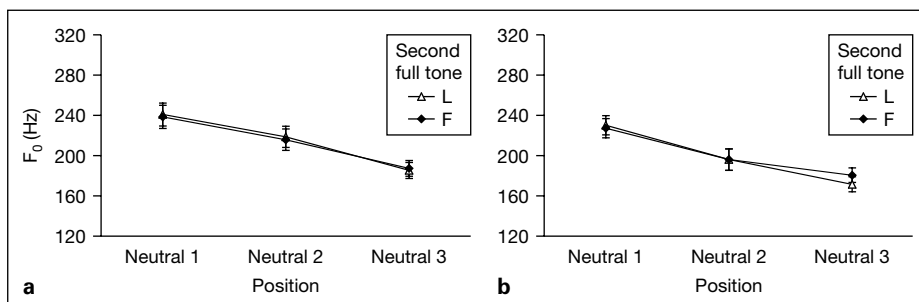
**Table 1.** Syllable duration (ms) as functions of speaking rate, focus condition and type of second tone

		On-focus		Prefocus	
		full	neutral	full	neutral
1st syllable	normal	266.7	219.5	238.0	198.2
	fast	204.6	174.3	179.0	158.4
2nd syllable	normal	213.9	128.6	209.4	126.0
	fast	183.9	113.8	178.6	113.3

**Fig. 5.** Mean  $F_0$  values at the middle (**a**) and end (**b**) of three consecutive neutral-tone syllables when the first full tone is H, R, L or F. The four lines show the mean  $F_0$  values when the first full tone varies across H, R, L and F. The leftmost points are mean  $F_0$  values at the end of the first full tones. They are plotted to show the course of  $F_0$  trajectory between the end of the first full tone to the middle and the end of the neutral tone.

The preceding tone has an apparent effect on the  $F_0$  contours of the neutral-tone syllables, as can be seen in figure 5. Specifically, the  $F_0$  of the first neutral tone is much influenced by the offset  $F_0$  of the preceding full tone (which is also the onset  $F_0$  of the neutral tone). As time elapses, i.e., from the onset to the middle (fig. 5a), and to the end (fig. 5b) of the first neutral tone, and from the first to the third neutral tone, the influence from the first full tone is gradually reduced. Two 3-factor repeated-measures ANOVAs were performed with the mid and end point  $F_0$  values as the dependent variable, and the preceding full tone (H/R/L/F), the second full tone (L/F), and the position of the neutral tone (1st/2nd/3rd) as independent variables. Results showed that  $F_0$  at both the middle and the end of the neutral-tone syllables was significantly influenced by (a) the first full tone [mid point:  $F(3, 9) = 6.46, p = 0.0127$ ; end point:  $F(3, 9) = 8.773, p = 0.0049$ ], and (b) the position of the neutral tone in the neutral-tone sequence [mid point:  $F(2, 6) = 37.67, p = 0.0004$ ; end point:  $F(2, 6) = 49.54, p < 0.0002$ ]. There were also significant interactions between the first full tone and the position at both the middle and the end of the neutral-tone syllables [mid point:  $F(2, 6) = 16.09, p < 0.0001$ ; end point:  $F(6, 18) = 9.797, p < 0.0001$ ].

Figure 5 also shows that the differences due to the first full tone are visible even by the middle and the end of the third neutral-tone syllable. Two 2-factor (first full tone,



**Fig. 6.** Mean  $F_0$  values at the middle (**a**) and the end (**b**) of three consecutive neutral-tone syllables when the second full tone is either L or F.

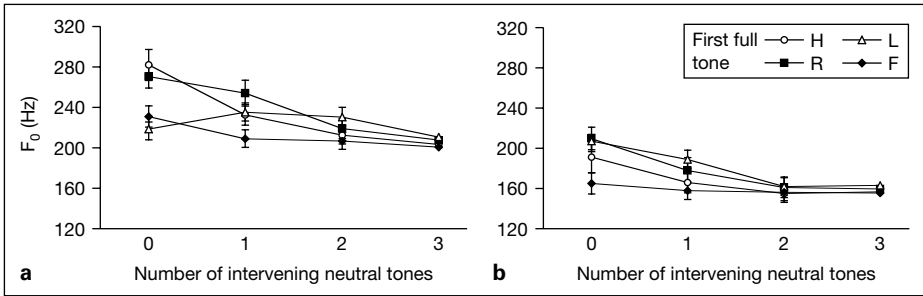
second full tone) repeated-measures ANOVAs were conducted on  $F_0$  at the middle and the end of the last of the three neutral-tone syllables. At the middle of the last neutral tone, the effect of the first full tone was highly significant [ $F(3, 9) = 39.239$ ,  $p < 0.0001$ ], while the effect of the second full tone was nonsignificant. At the end of the last neutral tone, the effect of the first full tone was highly significant [ $F(3, 9) = 16.54$ ,  $p = 0.0005$ ], while the effect of the second full tone was only marginally significant [ $F(1, 3) = 24.692$ ,  $p = 0.0157$ ]. Note that the end of the third neutral-tone syllable is three syllables away from the first full-tone syllable, but it is virtually the onset of the second full tone by the conventional definition.<sup>5</sup>

It is clear that the second full tone does not have much effect on the  $F_0$  values of neutral tones, as demonstrated by figure 6 [which shows the mean  $F_0$  values at the middle (fig. 6a) and end (fig. 6b) of the neutral-tone syllables as a function of the second full tone]. It seems that the second full tone does not have much effect on the  $F_0$  values of neutral tones. Two 3-factor repeated-measures ANOVAs [independent variables: first full tone (H/R/L/F), second full tone (L/F), position (1st/2nd/3rd neutral tone); dependent variables:  $F_0$  at the middle and end of neutral-tone syllables] found no main effect of the second full tone on the  $F_0$  of the neutral tone at either the middle or end of syllable. There were significant interactions between the second full tone and the position on the  $F_0$  at the middle of neutral tones [ $F(6, 18) = 11.81$ ,  $p = 0.0083$ ] and at the end of the neutral tone [ $F(6, 18) = 68.66$ ,  $p \leq 0.0001$ ]. But as can be seen in figure 6, the effects were very small in magnitude. The largest effect was 9.1 Hz at the end of the last neutral-tone syllable.

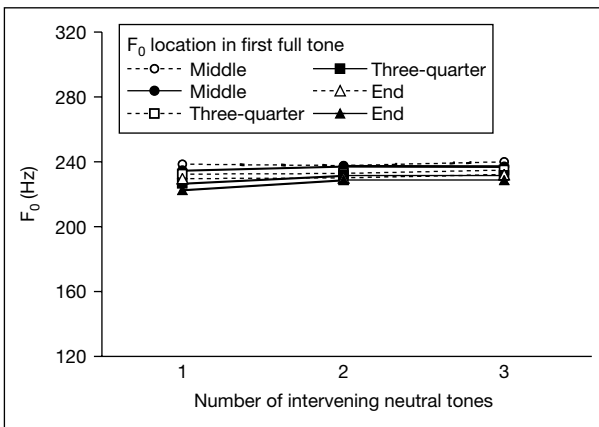
#### 4.2.3 $F_0$ Variations in the Full Tones

A consequence of the extensive influence of the first full tone on the following neutral tone(s) is that the same influence is still in effect by the onset of the second full tone. This can be seen clearly in figure 3. To see how far this influence penetrates into the second full tone, we plotted peak  $F_0$  of the second F tone and middle  $F_0$  of the second L tone in figure 7. In figure 7a, the difference in peak  $F_0$  in the second F tone due to the first full tone is visible even when there are two intervening neutral-tone

<sup>5</sup> Xu and Liu [2002] and Liu and Xu [2003] reported evidence suggesting that the boundary between any two adjacent syllables is actually earlier than the conventional definition. Applying this new understanding would make the asymmetry between carryover and anticipatory tonal effect even more dramatic.



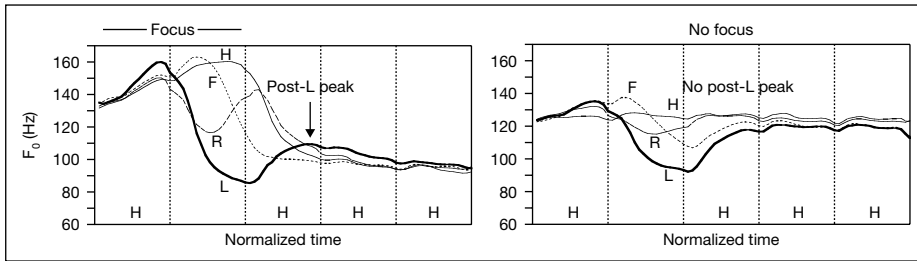
**Fig. 7.** Mean peak  $F_0$  (a) and mean  $F_0$  at the middle of the second full tone (b) when it is preceded by 0–3 neutral-tone syllables and when the first full tone is H, R, L or F.



**Fig. 8.** Mean  $F_0$  values at the middle, three-quarter and end of the first full tone when it is followed by 0–3 neutral-tone syllables and when the second full tone is either L (open symbol with dashed line) or F.

syllables. In figure 7b, the  $F_0$  difference in the middle of the second L tone is visible when there is one intervening neutral tone. Two 3-factor (first full tone, number of neutral tones and focus) repeated-measures ANOVAs were performed on the peak  $F_0$  of the second F tone and middle  $F_0$  of the second L tone. For the peak  $F_0$ , the three main effects were all significant [first full tone:  $F(3, 9) = 18.01, p = 0.0004$ ; number of neutral tones:  $F(3, 9) = 18.69, p = 0.0003$ ; focus:  $F(1, 3) = 12.29, p = 0.0393$ ]. There was a significant interaction between the first full tone and the number of neutral tones [ $F(9, 27) = 10.25, p < 0.0001$ ], indicating the reduction of the effect of the first full tone over time as the number of neutral tones increased. There was also a significant interaction between the first full tone and focus [ $F(3, 9) = 35.98, p < 0.0001$ ]. As can be seen in figure 3, focus on the target noun phrase lowers the  $F_0$  of the second full tone, but the effect is lessened when the first full tone is L due to the post-L  $F_0$  raising briefly discussed earlier (end of section 4.1).

Figure 8 illustrates the effects of the following tones on the  $F_0$  of the first full tone. Displayed here is the  $F_0$  at the middle, three-quarter and end of the first full-tone syllable when the second full tone is L (open symbol with dashed line) or F. The effects were quite



**Fig. 9.** Mean  $F_0$  contours five-syllable sentences adapted from Xu [1999]. Each curve is an average of 40 repetitions by 8 subjects.

small. Three 3-factor (first full tone, second full tone and number of intervening neutral tones) repeated-measures ANOVAs found the effects of the second full tone to be significant at all three locations in the first full tone [middle:  $F(1, 3) = 10.54, p = 0.0476$ ; three-quarter:  $F(1, 3) = 33.19, p = 0.0104$ ; end:  $F(1, 3) = 92.52, p = 0.0024$ ]. Interestingly, in each case, the  $F_0$  was higher when the second full tone was L than when it was F. Thus, this ‘anticipatory’ effect is dissimilatory and is therefore different in nature from the carryover difference which is large and assimilatory. Such asymmetry is similar to previous reports on the influences of adjacent full tones on each other (for Asian tonal languages: see Xu [1997, 1999] for Standard Chinese, and Gandour et al. [1994] for Thai; for African tonal languages: see Laniran and Clements [2003] and Connell and Ladd [1990] for Yoruba, Mountford [1983] for Bambara, and Hyman [1993] for Kirimi, as cited in Gussenhoven [2004]).

#### 4.2.4 Post-L $F_0$ Raising

As briefly discussed in section 4.1, when the first full tone was L, the following  $F_0$  rose throughout the first post-L neutral-tone syllable and the rise continued in much of the second neutral-tone syllable. This effect is also clearly visible in figure 5, and even in figure 7, where both peak  $F_0$  in the second F tone and middle  $F_0$  in the second L tone are often the highest when the first tone is L and when there are intervening neutral tone(s). The rising contour of the neutral tone immediately after the L tone has long been known [Chao, 1968; Lin and Yan, 1980; Shih, 1987], and it has been suggested to be related to the final rise of the L tone produced in isolation [Yip, 1980]. However, post-L raising has also been reported to occur in a full tone when the preceding L tone is focused [Chen, 2003; Xu, 1995, 1999]. Figure 9 shows that the  $F_0$  of the H tone is higher after a focused L tone than after other focused tones [data from Xu, 1999]. Although this raising can again be linked to the final rise of the L tone in isolation [Xu, 1995], it is also reminiscent of a post-L raising phenomenon in English. Pierrehumbert [1980] has reported that there is a constant time interval between  $F_0$  valley due to an L\* pitch accent in English and the following  $F_0$  peak. She finds that the time interval remains stable at 191–202 ms. To determine if the alignment of the post-L peak is also constant in Standard Chinese, we performed a simple linear regression analysis on the present data with the magnitude of the valley-to-peak excursion as the predictor and the time interval between the valley and the peak as dependent variable. There was no correlation between the two and the slope of the regression line was virtually zero. The mean time interval between the valley and peak was 232.276 ms, which was a bit

longer than that reported by Pierrehumbert [1980]. We also performed simple linear regression analysis on data from Xu [1999] in cases where the second tone was L in a 5-tone sequence and when it was under focus. The data were from 8 subjects, 4 males and 4 females, with a total of 10 utterances from each subject. Again, we found no significant relation between the magnitude of  $F_0$  excursion and time interval between the valley and the peak when male and female subjects were analyzed separately.<sup>6</sup> The mean time interval for that set of data was 175 ms.

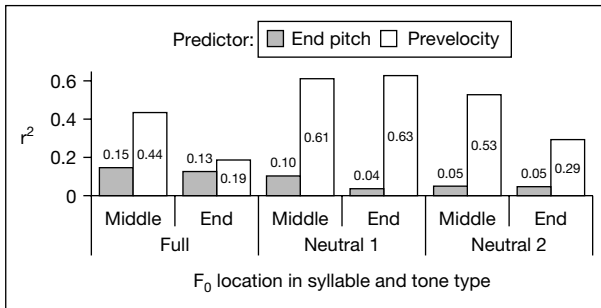
The difference in the alignment of post-L  $F_0$  peak between the present data and the data from Xu [1999] could be related to two factors. First, in Xu [1999], the tone after L is a full tone, whereas in the present data it is a neutral tone. Second, in Xu [1999], the post-L H is in a postfocus position, whereas in the present data, the neutral tone is either under focus or before the focus. Being in a postfocus position and a full tone, the H tone in Xu [1999] seems to be able to reverse the  $F_0$  raising induced by its preceding L sooner than the neutral tone in the present study.

#### 4.2.5 Sources of $F_0$ Variation in the Neutral Tone

The finding (in section 4.2.2) that the second full tone has no assimilatory influence on its preceding neutral tone indicates that the upcoming full tone is not responsible for the  $F_0$  of the neutral tone. The finding (also in section 4.2.2) that the first full tone has extensive influence on the following neutral-tone  $F_0$  indicates that the preceding tone is a major source of the variability in the neutral-tone  $F_0$ . Nevertheless, the clear converging trends in neutral tone(s) as seen in figures 3 and 5, which are in a direction away from the influence of the preceding full tone, indicate that a second source of  $F_0$  variability in the neutral tone exists, and it is likely to be a pitch target belonging to the neutral tone itself. This is because the converging pattern is similar in nature to those in the full tones, for which there is little doubt as to the existence of their tonal targets. Figures 3 and 5 and the analyses in section 4.2.2 also show, however, that the convergence of  $F_0$  contours in the neutral tone is not complete even by the end of the third neutral-tone syllable. This is in sharp contrast to the almost complete merger of the  $F_0$  contours in the second full tone by the end of the syllable. It thus appears that the difference between a full tone and a neutral tone lies mainly in how effectively they overcome the influence of the preceding tone while approaching their respective targets.

As discussed in section 4.1 and clearly observable in figures 3 and 5, the influence of the preceding tone is of two forms, the offset  $F_0$  and the final velocity. To estimate the magnitude of such an influence and how effective it is during the production of the full tone and the neutral tone, we performed simple linear regressions on the  $F_0$  at the middle and the end of the second full-tone syllable and the first and second neutral-tone syllables. The  $F_0$  values were measured in semitones with the offset  $F_0$  of the last syllable (*shuo*) in the carrier sentence as the reference. End pitch is the offset  $F_0$  of the first full-tone syllable. Prevelocity is the peak velocity in the last quarter of the first full-tone syllable, measured in semitones/second. The mean  $r^2$  values of the regression analyses are plotted in figure 10. Two consecutive neutral tones were examined because, as discussed in section 4.2.1, neutral tone syllables are about 61% the length of full-tone syllables. So, the end of a full-tone syllable is temporally close to the middle of the

<sup>6</sup> When data from male and female subjects were analyzed together, there was a weak correlation. But a closer examination found that the weak correlation was virtually totally due to the male/female difference, and the difference is mainly due to excessive creaky voice in the female data.



**Fig. 10.** Mean  $r^2$  values of simple regression analyses on  $F_0$  at the middle and end of the second full tone and first neutral tone, with end pitch, prevelocity and pitch  $\times$  velocity as predictors.

second neutral-tone syllable. Because of the special mechanism involved with the L tone (i.e. L tone sandhi in a sequence of two low tones), sentences in which the first full tone is L were excluded from these regression analyses.

Three patterns can be observed in figure 10. First, overall, end pitch is a much less effective predictor of  $F_0$  in the following tone than prevelocity. Second, prevelocity predicts the  $F_0$  of the neutral tone much better than that of full tone. Third, the influence of prevelocity on a full tone is much reduced by the end of the full-tone syllable, whereas there is no reduction by the end of the first neutral-tone syllable, and the prediction is still quite good at the middle or even the end of the second neutral tone. Both the second and third observations indicate that the neutral tone is much less effective than a full tone in overcoming the influence of the preceding tone. This suggests weaker articulatory strength applied during the implementation of a neutral tone than during that of a full tone.

Finally, to estimate the target value of the neutral tone, we computed the final  $F_0$  values of the third neutral-tone syllable in the three-neutral-tone sequences, and compared them to the average of the maximum  $F_0$  in the second F tone and minimum  $F_0$  in the second L tone (in semitones). The former was on average 1 semitone lower than the latter. A 3-factor repeated-measures ANOVA (independent variable: first full tone, focus and tone type) showed that the difference due to tone type was nonsignificant. This suggests that the targeted  $F_0$  for the neutral tone is about halfway between the highest and lowest  $F_0$  values of the full tones.<sup>7</sup>

## 5 Discussion

The data obtained in the present study demonstrate that the  $F_0$  contours of the neutral tone in Standard Chinese cannot be derived from tone spreading from the preceding tone or linear interpolation between adjacent full tones. As shown in figures 3 and 5,

<sup>7</sup> Note that this does not suggest that the neutral tone uses the following full tone as reference for the mid height. Rather, based on the finding that the following full tone has no influence on the preceding neutral tone, it should be the following full tone that takes the end point of the neutral tone as the reference point, and then rises or falls from there to realize the max  $F_0$  in F or min  $F_0$  in L.

the  $F_0$  of a neutral-tone syllable drops between an H and an F tone, which is not predicted by tone spreading or linear interpolation. The finding in section 4.2.2 that a full tone has no assimilatory effect on a preceding neutral tone makes interpolation even more unlikely, because interpolation would have generated a symmetrical influence from the flanking full tones. The finding in section 4.2.3 that the peak  $F_0$  or  $F_0$  at the mid point of the full tone following the neutral tone is influenced both by the first full tone (before the neutral tone) and by the number of neutral tones, as shown in figures 4 and 7, further demonstrates that it is unlikely that the  $F_0$  of the neutral tone comes from interpolation between the preceding full tone and an upcoming boundary tone with a fixed  $F_0$  value. This is because such an interpolation would have involved a process that is circular: the end point of interpolation is dependent on the preceding neutral tone(s) whose  $F_0$  is in turn dependent on the interpolation.

Our data analysis instead reveals two critical characteristics of the neutral tone. First, the  $F_0$  of a string of neutral-tone syllables is approaching some kind of underlying target whose identity becomes increasingly evident as the number of consecutive neutral tones increases. The existence of such a target is indicated by the gradual convergence of  $F_0$  contours over time toward a mid value, in spite of the differential influences of the preceding full lexical tones. Second, whatever the value of this underlying target may be, it is implemented quite differently from the full lexical tones: it is realized rather sluggishly and the sluggishness leads to greater and more sustained influence from the preceding tone.<sup>8</sup>

One may argue that what these results have shown is no more than the fact that there is greater variability in the neutral tone than in the full lexical tones in Standard Chinese, and that such variability is just further evidence that the neutral tone does not have the same status as a full lexical tone. Such an argument, however, would not help us understand where the wide range of variations in the neutral-tone contours come from, and why, despite the variability, the pitch values do tend to converge over time. One may further question that if the neutral tone has a target, why this target is not implemented the same way as a lexical tonal target. As shown in figure 3, the approximation of a lexical full-tone target is mostly completed by the end of its host syllable. In contrast, figures 3 and 5 show that the approximation of the neutral-tone target is not fully completed even by the end of the third neutral-tone syllable. Furthermore, the results of regression analyses indicate that the variations in the  $F_0$  of the neutral-tone are due mostly to the first full tone immediately preceding the neutral-tone sequence. Such an influence is exerted through both the final  $F_0$  height and the final velocity of the first full tone. Any theory about the variability of the neutral tone therefore has to be able to explain these results.

There are many proposed accounts for contextual variations [see Hardcastle and Hewlett, 1999, for detailed reviews], but three have been especially influential. One account, known as articulatory phonology, attributes the variations to the temporal overlap (or coproduction) of articulatory gestures [see e.g. Browman and Goldstein, 1992, and references therein]. Based on this framework, one possibility is to attribute the variations observed in the present study to the overlap of the neutral tone with the surrounding tones. However, as our data have shown, first, the following full tone has

<sup>8</sup> As pointed out by Moira Yip [personal commun.], if the variability of the neutral tone is length related, one might expect to see similar variability in other very short tones, such as Cantonese Ru-sheng syllables, which does not happen.

no influence on the preceding neutral tone. So the neutral tone is unlikely to be overlapped by the following tone. Second, gestures are assumed to have invariant durations because their timing is intrinsic [Fowler, 1980]. The temporal scope of the influence of the preceding full tone on the following neutral tone, however, varies extensively with the number of consecutive neutral tones. An overlap account would thus be incompatible with the intrinsic timing assumption. Another account, first explicitly articulated by Keating [1988b, 1990] and known as the ‘window theory’, assumes that phonological units exhibiting a large amount of variation are unspecified both phonologically and phonetically. The surface values of these unspecified units are described as due to a wide window which allows direct interpolation between the preceding and the following units. Along a similar line, but in a more recent incarnation, is the exemplar theory [e.g. Pierrehumbert, 2002], which would attribute the variations to the cluster of representations of the same phonological unit with varied phonetic details. Note that theories like these are useful only if our sole task is to tally the amount of variability. They cannot explain the detailed variation patterns in the present data, including the asymmetrical influences from the surrounding tones, the gradual convergence over time and the sensitivity to final velocity of the preceding tone. The third account is the hyper- and hypo-articulation (H&H) hypothesis proposed by Lindblom [1990], which explains contextual variation as the consequence of a continuous adaptation of speech production to the changing demands of communication. According to the H&H hypothesis, the amount of coarticulation is inversely related to the demand for contrast. Thus, the greater variability of the neutral tone should be due to a weaker demand for contrasting it with other tones. However, phonemically, the neutral tone *is* different from all other tones in Standard Chinese, as explained in the Introduction. Therefore, the differences found in the present study should be viewed as *helpful* for contrasting the neutral tone with other tones, rather than making it less distinguishable from them.

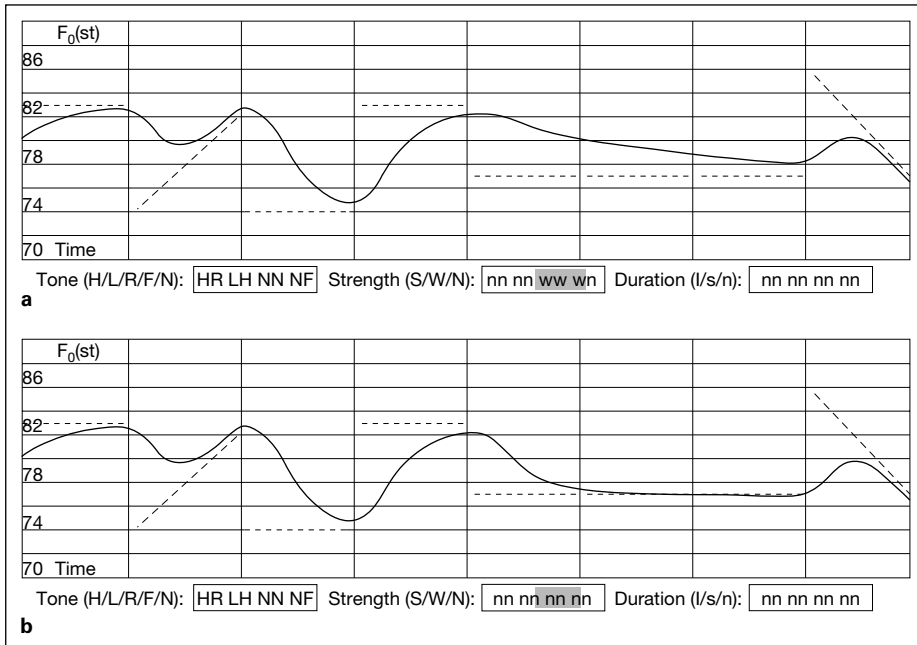
A more coherent account of the present data, we would like to suggest, is to posit that the neutral tone does have a specific underlying pitch target, which is likely to be static and mid, based on the analysis results presented in section 4.2.5. But in addition, this target is likely to be implemented with a weak articulatory strength, as suggested by the regression analysis shown in figure 10. It is this weak strength that seems to give rise to the large surface variability in the neutral tone.<sup>9</sup> Given a mid pitch target and a weak articulatory strength, the implementation of the neutral tone is probably not different from that of the full tones in Standard Chinese, i.e., through a process of *syllable-synchronized sequential target approximation*, summarized as the target approximation model [Xu, 2005; Xu and Wang, 2001]. Under this model, to produce any tone, the speaker has to have a specific articulatory goal, referred to as a pitch target, which can be specified by a simple linear function [Xu et al., 1999]. The task of the speaker is then to articulatorily implement this target within the time interval allocated to it, which results in an asymptotic approximation of the target. In this model, the speed at which the target is approximated is specifiable, as it is determined by the amount of articulatory strength applied to the implementation of the target. Based on the present data, the neutral tone is implemented with a much weaker force than the other tones. More specifically, implementing a target with a strong force results in faster approximation of the target as well as faster relief from the influence of the preceding tone.

<sup>9</sup> It is even possible that this weak strength is part of the underlying specification of the neutral tone. We leave such a possibility, however, to further research.

Implementing a target with a weak force, in contrast, would result in slower approximation of the target as well as slower relief from the influence of the preceding tone. This mechanism thus gives rise to the large influence of the preceding lexical tone over the neutral tone, and it explains why it takes a span of several syllables for the target of the neutral tone to emerge. This also explains why previous work has reported that the pitch contour of the neutral tone is solely determined by the preceding tone. It is because the resistance of the neutral tone against the influence of the preceding tone is so small that it is only through detailed analysis of the time course of  $F_0$  contours of the neutral tone as well as careful comparisons with that of full tones that the target of the neutral tone can be revealed.

This account shares some of the characteristics of the recently proposed soft template (or Stem-ML) model of intonation [Kochanski and Shih, 2003], which has been applied to Standard Chinese (though not yet to the neutral tone) in Kochanski et al. [2003]. The soft template model [Kochanski and Shih, 2003] describes  $F_0$  contours as resulting from implementing underlying tonal templates with different amounts of muscle forces under the physical constraint of *smoothness*. The smoothness constraint makes the connection between adjacent templates seamless, and the varying muscle forces determine the degree to which the shape of each template is preserved in the surface  $F_0$ . While in both models articulatory strength accounts for the behavior of the neutral tone, one property of the soft template model contrasts critically with the target approximation model. That is, the smoothness constraint in the soft template model is *bidirectional*, i.e., exerting both carryover and anticipatory assimilatory influences, while the mechanism of target approximation only allows the preceding tone to influence the following tone, but not vice versa. The finding of the present study that the following full tone has virtually no influence on the preceding neutral tone is thus more consistent with the target approximation model than with the soft template model.

Finally, there have been suggestions that the surface  $F_0$  of the neutral tone results from relaxation to a rest position [see discussion in Yip, 2002]. Relaxation toward a rest position is a core component of the command-response model [Fujisaki, 1983], in which  $F_0$  automatically returns to a baseline when there are no active muscle commands. What is critical about such relaxation accounts is that they assume that the vocal folds are stretched beyond their resting length during phonation. As discovered by Hollien [1960] and Hollien and Moore [1960], however, the vocal folds are typically much shorter during phonation than during rest. Thus, any  $F_0$  drop is likely to be achieved by active contraction of the pitch-lowering muscles such as the thyroarytenoids (vocalis) in conjunction with the relaxation of the pitch-raising muscle such as the cricothyroids. There is therefore no 'automatic' return to a rest position. Besides Standard Chinese, there are other Mandarin languages which also have the category neutral tone but different lexical tonal inventories from Standard Chinese. More detailed studies on the  $F_0$  pattern of neutral tone in these languages could be illuminating. If indeed these languages have neutral-tone targets specified with different  $F_0$  values from that in Standard Chinese, it would support our proposal that the neutral tone has a phonologically specified tonal target. Even if all dialects seem to have a similar  $F_0$  pattern, we wish to point out that whether  $F_0$  automatically returns to a neutral level should be treated differently from whether there is a default neutral pitch register. It is quite plausible that there is a neutral pitch register near the level of the habitual pitch [Zemlin, 1988], which is used as a phonological specification for weak tones such as the neutral tone in Mandarin. However, such pitch register is likely a real target that has



**Fig. 11.** Simulation of the effect of articulatory strength with a preliminary quantitative implementation of the target approximation model. Vertical lines: syllable boundaries; dashed lines: underlying pitch targets (defined by  $y = b + ax$ , where  $b$  is  $y$ -intercept and  $a$  is slope), and solid curve: surface  $F_0$  trajectory. The labels below each graph indicate for each syllable its tone (H, R, L, F), strength ( $s$  = strong,  $w$  = weak,  $n$  = normal), and duration ( $l$  = long,  $s$  = short,  $n$  = normal). In the top graph, the strength of syllables 5–7 is set to weak, resulting in slow approximation of the static pitch targets. In the bottom graph, the strength of these syllables is set to normal, resulting in faster approximation of the pitch targets.

to be implemented with active muscle force, just as the other pitch targets do, although the strength of the muscle force may be weak.

An additional support for the mid-static target plus weak strength account comes from our preliminary quantitative modeling effort, as shown in figure 11, which shows the displays of an interactive Java applet implementation of the target approximation model ([www.phon.ucl.ac.uk/home/yi/f0\\_model.html](http://www.phon.ucl.ac.uk/home/yi/f0_model.html)). In figure 11a,  $F_0$  of syllables 5–7 (syllable boundaries are indicated by vertical lines) is generated with weak strength, which results in surface  $F_0$  contours that approach the consecutive static targets rather slowly, resembling the neutral-tone contours in figure 3. In contrast, in figure 11b, the strength of syllables 5–7 is set to normal, resulting in quick approximation of the targets, and the  $F_0$  contours are quite uncharacteristic of those shown in figure 3. Though still quite unsophisticated, this preliminary implementation of the target approximation model can nevertheless simulate the main characteristics of the neutral tone using static targets and weak strength.

Such an articulation-oriented account of the neutral tone is potentially applicable to similar phenomena in other languages. The  $F_0$  contour of neutral tone, in the particular context of being preceded by an H tone and followed by an F tone, exhibits a very similar pattern as the transition between two  $H^*$  tones in English described by Ladd

and Schepman [2003]. With an increasing number of neutral-tone syllables, there appears a steeper declining  $F_0$  contour between the two lexical tones. Comparably, with an increasing number of segments between two  $H^*$  pitch accents, there is a steeper declining  $F_0$  contour. Ladd and Schepman conclude with the existence of an L target in the second  $H^*$  and propose that the difference between such an L target and that in the Pierrehumbert L +  $H^*$  sequence may be due to the different phonetic realizations of the same phonological L target in English. Given what we have observed in Mandarin neutral tone, we may speculate that the L in that particular case could be similar in nature as the neutral tone in Standard Chinese, i.e., it is a mid target that is implemented with a weak force. This possibility has been investigated by Xu and Xu [2005], and patterns similar to those reported in the present study have been found.

As discussed in section 4.2.4, not all  $F_0$  variations in the neutral tone can be explained by implementation of a mid target under the mechanical influence of the preceding tone. When the first full tone is L, the  $F_0$  during the L continues to drop till the end of the syllable. The  $F_0$  of the following neutral-tone syllable, however, rises throughout its duration and the rise continues in much of the second neutral-tone syllable. As a result, the  $F_0$  of the second and third post-L neutral-tone syllables is higher than that after the other three full tones. The post-L  $F_0$  rise in a neutral tone has been suggested to be due to a floating H in an L tone which surfaces only when the L tone is in isolation (thus giving it the well-known dipping contour) or in the post-L neutral tone [Yip, 1980]. However, our data analyses in section 4.2.4 suggest a more universal source for such a post-L bounce. A parallel is found between the rather constant valley-to-peak time interval in the L-neutral sequences in the present data, in the LH sequence in Standard Chinese when the L is focused, based on reanalyzed data from Xu [1999], and in the post-valley  $F_0$  peak after an  $L^*$  accent in English, as reported by Pierrehumbert [1980]. Thus, there seems to be a tendency for  $F_0$  to ‘bounce’ back after reaching an extremely low value. The mechanism of such a bounce can only be speculated on. It has been well established that the production of very low  $F_0$  involves not only changes in the activities of the cricothyroid and the thyroarytenoid muscles, but also the activation of the strap muscles, including the sternohyoids, omohyoids and sternothyroids [Erickson, 1976; Erickson et al., 1995; Hallé, 1994; Fujisaki, 2003]. The sudden disengagement of these powerful muscles after the implementation of a [low] pitch target may have created a temporary imbalance among the muscle forces that jointly control the vocal fold tension, thus resulting in a momentary overcorrection. While the exact nature of the post-L bounce can be revealed only in future studies, our data analyses at least suggest a mechanism that is different from (though, of course, also related to) other  $F_0$  control mechanisms.

Finally, the findings about the neutral tone are also relevant for the understanding of speech production at the segmental level. A closely related case is the schwa in English. The surface realization of the schwa has been reported to exhibit very robust contextual variations. Two main proposals have been put forward: one is that the schwa is underspecified not only at the phonological level but also at the phonetic level, which gives rise to a wide ‘window’ for feature interpolation between its preceding and following targets [Keating, 1990]. Another is that the schwa does have an articulatory target and it is the blending of the schwa with the articulation of the following vowel that gives rise to the observed variations [Browman and Goldstein, 1992]. Given what we have observed in the neutral tone, it would be interesting to ask whether the articulatory target of schwa would also emerge when uttered in a sequence of schwa targets. Or we

may also slow down the production of such sequences and see when given more time, whether the schwa target will become apparent. This methodology has been adopted in a study on lip protrusion and velum movement in Boyce et al. [1992] (as mentioned earlier). More research is certainly needed to gain greater insight into the issue of contextual variability in speech. The data on the neutral tone in the present study would serve only as a starting point.

## 6 Conclusion

Previous studies have reported extensively on the large amount of variability in the  $F_0$  contours of the neutral tone in Standard Chinese, most of which have further linked such variability to the tone that precedes the neutral tone. The general consensus among these studies, quite logical as it first seems, is that the neutral tone simply does not carry any phonological specifications, nor does it have any specific phonetic target. The surface  $F_0$  contours of the neutral tone are assumed to be derived from mechanisms such as spreading or interpolation. The detailed acoustic analyses performed in the present study confirm (a) increased variability in the neutral tone as compared to the full lexical tones and (b) the preceding tone as the main source of such variability. More importantly, our data also show that increased variability does not necessarily mean lack of underlying target. On the contrary, our data suggest that to account for the detailed  $F_0$  contours of the neutral tone in different tonal contexts in a coherent manner, it is best to assume that the neutral tone does have a target, but at the same time the target is implemented with robustly reduced effort as compared to the full lexical tones. Such an account can explain (1) why the variability in the neutral tone reduces over time, (2) why the direction and magnitude of the  $F_0$  contour variations in the neutral tone is largely predictable by the final velocity of the preceding tone, (3) why the  $F_0$  contour would start to turn away from the direction of the influence of the preceding tone even within the first neutral-tone syllable following that tone, and yet the remnant of the influence is still visible by the end of the third neutral tone in a row, and, finally, (4) why the  $F_0$  contour of a full lexical tone following a neutral tone takes the  $F_0$  offset of the neutral tone as its starting point just as it takes the  $F_0$  offset of another full tone as its starting point. This account of the neutral tone in Standard Chinese strengthens the model of tone production for full lexical tones [Xu and Wang, 2001], but also emphasizes the importance of the strength of tonal implementation. Much more research, however, remains to be done to see if the principles that our new account alludes to are applicable to other languages and to aspects of speech other than tone. Finally, the link between the post-L bounce in the neutral tone and that in the full Mandarin tones as well as in the post-L peak in English points to a possible independent mechanism that also calls for further investigation.

## Appendix: A Full List of Materials Elicited in this Experiment

### A Question-Answer Pairs to Illustrate How Focus Was Elicited on Different Parts of the Utterance (Focused Constituent in Bold)

#### (1) Target word under focus

Question on subject

*tā shuō shéi měi duō le?*

He said who beautiful more aspect marker

‘Who did he say is more beautiful?’

Answer

*tā shuō **mā** měi duō le.*

He said mother beautiful more aspect marker

‘He said the mother is more beautiful.’

#### (2) Target word without focus

Question on predicate

*tā shuō mā zěnmé le?*

He said mother what aspect marker

‘What did he say about the mother?’

Answer

*tā shuō mā **měi duō** le.*

He said mother beautiful more aspect marker

‘He said the mother is **more beautiful**.’

### B List of Utterances Elicited (Target Constituents Bold)

Group A: neutral-tone sequence preceded by a high tone (*mā*) and followed by either a falling (*màn*) or a low (*měi*) tone.

#### (1) No neutral tone

*tā shuō **mā** màn/měi duō le.*

‘He said that the mother is slower/more beautiful.’

#### (2) One neutral tone

*tā shuō **māma** màn/měi duō le.*

‘He said that the mother is slower/more beautiful.’

#### (3) Two neutral tones

*tā shuō **māmamen** màn/měi duō le.*

‘He said that the mothers are slower/more beautiful.’

#### (4) Three neutral tones

*tā shuō **māmamende** màn/měi duō le.*

‘He said that what the mothers have is slower/more beautiful.’

Group B: neutral-tone sequence preceded by a rising tone (*miáo* or *maó*) and followed by either a falling (*màn*) or a low (*měi*) tone.

#### (1) No neutral tone

*tā shuō **miáo** màn/měi duō le.*

‘He said that the seed (grows) more slowly/is more beautiful.’

#### (2) One neutral tone

*tā shuō **maomao** màn/měi duō le.*

‘He said that maomao (name sg.) is slower/more beautiful.’

#### (3) Two neutral tones

*tā shuō **maomaomen** màn/měi duō le.*

‘He said that maomao (name pl.) are slower/more beautiful.’

- (4) Three neutral tones  
*tā shuō máomaomende màn/měi duō le.*  
 ‘He said that what maomao (name pl.) have is slower/more beautiful.’

Group C: neutral-tone sequence preceded by an L tone (*niǎo* or *nǎi*) and followed by either a falling (*màn*) or a low (*měi*) tone.

- (1) No neutral tone  
*tā shuō niǎo màn/měi duō le.*  
 ‘He said that the bird is slower/more beautiful.’
- (2) One neutral tone  
*tā shuō nǎinai màn/měi duō le.*  
 ‘He said that the grandma is slower/more beautiful.’
- (3) Two neutral tones  
*tā shuō nǎinaimen màn/měi duō le.*  
 ‘He said that grandmas are slower/more beautiful.’
- (4) Three neutral tones  
*tā shuō nǎinaimende màn/měi duō le.*  
 ‘He said that what the grandmas have is slower/more beautiful.’

Group D: neutral-tone sequence preceded by a falling tone (*mèi*) and followed by either a falling (*màn*) or a low (*měi*) tone.

- (1) No neutral tone  
*tā shuō mèi màn/měi duō le.*  
 ‘He said that the sister is slower/more beautiful.’
- (2) One neutral tone  
*tā shuō mèimeī màn/měi duō le.*  
 ‘He said that the sister is slower/more beautiful.’
- (3) Two neutral tones  
*tā shuō mèimeimen màn/měi duō le.*  
 ‘He said that the sisters are slower/more beautiful.’
- (4) Three neutral tones  
*tā shuō mèimeimende màn/měi duō le.*  
 ‘He said that what the sisters have is slower/more beautiful.’

Note: most kinship terms in Standard Chinese can be monosyllabic or bisyllabic (with the second part reduplicated and carrying a neutral tone).

## Acknowledgements

We thank Ellen Broselow, Carlos Gussenhoven, Marie Huffman, Yoonjung Kang, and Moira Yip for insightful questions on earlier drafts of this paper. Our thanks also go to Xuejing Sun and Dongning Mao who provided critical technical help for our data analysis. Finally, we thank Klaus Kohler, Chilin Shih, and an anonymous reviewer for very helpful comments. This research was supported in part by the Moray Endowment Fund from the University of Edinburgh and the VENI Innovational Research grant from the Netherlands Organization for Scientific Research (NWO) to the first author, and by NIH grant No. DC03902 to the second author.

## References

- Aufterbeck, M.: Aspects of prehead and onset: the onset onglide phenomenon. Proc. 1st Int. Conf. Speech Prosody, Aix-en-Provence 2002, pp. 155–158.
- Boyce, S.E.; Krakow, R.A.; Bell-Berti, F.: Phonological underspecification and speech motor organization. *Phonology* 8: 210–236 (1992).
- Browman, C.P.; Goldstein, L.: Targetless schwa: an articulatory analysis; in Ladd, Papers in laboratory phonology II: gesture, segment, prosody, pp. 26–36 (Cambridge University Press, Cambridge 1992).

- Bruce, G.: Swedish word accents in sentence perspective; in Malmberg, Hadding, *Travaux de l'Institut de Linguistique de Lund*, vol. 12 (Gleerup, Lund 1977).
- Chao, Y.R.: *A grammar of spoken Chinese* (University of California Press, Berkeley 1968).
- Chen, M.Y.: *Tone sandhi: patterns across Chinese dialects* (Cambridge University Press, Cambridge 2000).
- Chen, Y.: *The phonetics and phonology of contrastive focus in standard Chinese*; PhD diss. State University of New York, Stony Brook, N.Y. (2003).
- Cheng, C.C.: *A synchronic phonology of Mandarin Chinese* (Mouton, Paris 1973).
- Choi, J.D.: An acoustic-phonetic underspecification account of Marshalese vowel allophony. *J. Phonet.* 23: 323–347 (1995).
- Cohn, A.: Nasalisation in English: phonology or phonetics. *Phonology* 10: 43–83 (1993).
- Cruttenden, A.: *Intonation* (Cambridge University Press, Cambridge 1997).
- Crystal, D.: *Prosodic systems and intonation in English* (Cambridge University Press, Cambridge 1969).
- D'Imperio, M.: Tonal structure and pitch targets in Italian focus constituents. *Proc. 14th Int. Congr. Phonet. Sci.*, San Francisco 1999, vol. 3, pp. 1757–1760.
- Erickson, D.M.: *A physiological analysis of the tones of Thai*; PhD diss. University of Connecticut (1976).
- Erickson, D.; Honda, K.; Hirai, H.; Beckman, M.E.: The production of low tones in English intonation. *J. Phonet.* 23: 179–188 (1995).
- Face, T.: Focus and early peak alignment in Spanish intonation. *Probus* 13: 223–246 (2001).
- Fowler, C.A.: Coarticulation and theories of extrinsic timing. *J. Phonet.* 8: 113–133 (1980).
- Fujisaki, H.: Dynamic characteristics of voice fundamental frequency in speech and singing; in P.F. MacNeilage, *The Production of Speech*, pp. 39–55 (Springer-Verlag, New York 1983).
- Fujisaki, H.: Prosody, information, and modeling – with emphasis on tonal features of speech. *Proc. Workshop on Spoken Lang. Processing, Mumbai 2003*, pp. 5–14.
- Gandour, J.; Potisuk, S.; Dechongkit, S.: Tonal coarticulation in Thai. *J. Phonet.* 22: 477–492 (1994).
- Gick, B.: An X-ray investigation of pharyngeal constriction in American English schwa. *Phonetica* 59: 38–48 (2002).
- Gussenhoven, C.: *A semantic analysis of the nuclear tones of English*. (Indiana University Linguistics Club, Bloomington 1983).
- Gussenhoven, C.: *The phonology of tone and intonation* (Cambridge University Press, Cambridge 2004).
- Hallé, P.A.: Evidence for tone-specific activity of the sternohyoid muscle in modern Standard Chinese. *Lang. Speech* 37: 103–123 (1994).
- Hardcastle, W.J.; Hewlett, N.: *Coarticulation: theory, data and techniques* (Cambridge University Press, Cambridge 1999).
- Hollien, H.: Vocal pitch variation related to changes in vocal fold length. *J. Speech Hear. Res.* 3: 150–156 (1960).
- Hollien, H.; Moore, G.P.: Measurements of the vocal folds during changes in pitch. *J. Speech Hear. Res.* 3: 157–165 (1960).
- Huffman, M.K.: Phonetic patterns of nasalization and implications for feature specification; in Huffman, Krakow, *Phonetics and phonology 5: nasals, nasalization and the velum*, pp. 303–326 (Academic Press, San Diego 1993).
- Hyman, L.: Register tones and tonal geometry; in van der Hulst, Snider, *The phonology of tone: the representation of tonal register*, pp. 75–108 (Mouton de Gruyter, Berlin 1993).
- Jin, S.: *An acoustic study of sentence stress in Mandarin Chinese*; PhD diss. Ohio State University (1996).
- Keating, P.A.: The phonology-phonetics interface; in Newmeyer, *Linguistics: the Cambridge survey*. Volume 1. *Grammatical theory*, pp. 281–302 (Cambridge University Press, Cambridge 1988a).
- Keating, P.A.: Underspecification in phonetics. *Phonology* 5: 275–292 (1988b).
- Keating, P.A.: The window model of coarticulation: articulatory evidence; in Kingston, Beckman, *Papers in laboratory phonology. Part 1. Between the grammar and physics of speech*, pp. 451–470 (Cambridge University Press, Cambridge 1990).
- Kochanski, G.; Shih, C.: Prosody modeling with soft templates. *Speech Commun.* 39: 311–352 (2003).
- Kochanski, G.; Shih, C.; Jing, H.: Hierarchical structure and word strength prediction of mandarin prosody. *Int. J. Speech Technol.* 6: 33–43 (2003).
- Kohler, K.J.: Modelling prosody in spontaneous speech; in Sagisaka, Campbell, Higuchi, *Computing prosody*, pp. 187–210 (Springer, New York 1997).
- Kohler, K.J.: Prosody revisited – function, time, and the listener in intonational phonology. *Proc. Int. Conf. Speech Prosody, Nara 2004*, pp. 171–174.
- Ladd, D.R.; Schepman, A.: 'Sagging transitions' between high pitch accents in English: experimental evidence. *J. Phonet.* 31: 81–112 (2003).
- Laniran, Y.O.; Clements, G.N.: Downstep and high raising: interacting factors in Yoruba tone production. *J. Phonet.* 31: 203–250 (2003).
- Li, Y.J.; Lee, T.: Acoustical F<sub>0</sub> analysis of continuous Cantonese speech. *Proc. Int. Symp. Chinese Spoken Lang. Processing, Taipei 2002*, pp. 127–130.
- Li, Z.: *The phonetics and phonology of tone mapping in a constraint-based approach*; PhD diss. MIT, Cambridge, Mass. (2003).
- Lin, H.: *A grammar of Mandarin Chinese* (Lincom Europa, München 2001).
- Lin, M.; Yan, J.: *Beijingshua qingsheng de shengxue xingzhi. Dialect 3*: 166–178 (1980).

- Lindblom, B.: Explaining phonetic variation: a sketch of the H&H theory; in Hardcastle, Marchal, *Speech production and speech modeling*, pp. 413–415 (Kluwer, Dordrecht 1990).
- Liu, F.; Xu, Y.: Underlying targets of initial glides – evidence from focus-related  $F_0$  alignments in English. *Proc. 15th Int. Congr. Phonet. Sci.*, Barcelona 2003, pp. 1887–1890.
- Mountford, K.: *Bambara declarative sentence intonation*; PhD diss. Indiana University (1983).
- Myers, S.: Surface underspecification of tone in Chichewa. *Phonology 15*: 367–392 (1999).
- O'Connor, J.D.; Arnold, G.F.: *Intonation of colloquial English* (Longmans, London 1961).
- Pierrehumbert, J.: *The phonology and phonetics of English intonation*; PhD diss. MIT, Cambridge, Mass. (1980).
- Pierrehumbert, J.: Synthesizing intonation. *J. acoust. Soc. Am. 70*: 985–995 (1981).
- Pierrehumbert, J.; Beckman, M.: *Japanese tone structure* (MIT Press, Cambridge 1988).
- Pierrehumbert, J.: Word-specific phonetics; in Gussenhoven, Warner, *Laboratory phonology VII*, pp. 101–140 (Mouton de Gruyter, Berlin 2002).
- Shen, X.N.: Mandarin neutral tone revisited. *Acta Linguist. Hafniensia 24*: 131–151 (1992).
- Shih, C.: *The phonetics of the Chinese tonal system* (AT&T Bell Labs, Murray Hill 1987).
- van Santen, J.; Shih, C.; Möbius, B.: Intonation; in Sproat, *Multilingual text-to-speech synthesis: the Bell Labs approach*, pp. 141–190 (Kluwer Academic Publishers, Dordrecht 1998).
- Wang, J.: The representation of the neutral tone in Chinese Putonghua; in Wang, Smith, *Studies in Chinese phonology*, pp. 157–183 (Mouton de Gruyter, Berlin 1997).
- Warner, N.; Jongman, A.; Cutler, A.; Mücke, D.: The phonological status of Dutch epenthetic schwa. *Phonology 18*: 387–420 (2001).
- Xu, C.X.; Xu, Y.; Luo, L.-S.: A pitch target approximation model for  $F_0$  contours in Mandarin. *Proc. 14th Int. Congr. Phonet. Sci.*, San Francisco 1999, pp. 2359–2362.
- Xu, Y.: Contextual tonal variations in Mandarin. *J. Phonet. 25*: 61–83 (1997).
- Xu, Y.: Effects of tone and focus on the formation and alignment of  $F_0$  contours. *J. Phonet. 27*: 55–105 (1999).
- Xu, Y.: Speech melody as articulatorily implemented communicative functions. *Speech Commun. 46*: 220–251 (2005).
- Xu, Y.; Liu, F.: Segmentation of glides with tonal alignment as reference. *Proc. 7th Int. Conf. Spoken Lang. Processing*, Denver 2002, pp. 1093–1096.
- Xu, Y.; Sun, X.: Maximum speed of pitch change and how it may relate to speech. *J. acoust. Soc. Am. 111*: 1399–1413 (2002).
- Xu, Y.; Wang, Q.E.: Pitch targets and their realization: evidence from Mandarin Chinese. *Speech Commun. 33*: 319–337 (2001).
- Xu, Y.; Xu, C.X.: Phonetic realization of focus in English declarative intonation. *J. Phonet. 33*: 159–197 (2005).
- Yip, M.: *The tonal phonology of Chinese*; PhD diss. MIT, Cambridge (1980).
- Yip, M.: *Tone* (Cambridge University Press, Cambridge 2002).
- Zemlin, W.R.: *Speech and hearing science – anatomy and physiology* (Prentice Hall, Englewood Cliffs 1988).