In Proceedings of International Symposium on Tonal Aspects of Languages, 2006, La Rochelle, France: 3-13.

Principles of Tone Research

Yi Xu

Department of Phonetics and Linguistics
University College London, UK
yi@phon.ucl.ac.uk

Abstract

Current advances in tone research are rather uneven. The major obstacles to faster progress in the area are not in the lack of technological means, but in the mindset of our discipline. This paper discusses ways to improve tone research by considering a set of basic principles both in research methodology and in theoretical thinking.

1. Introduction

Modern scientific study of tone goes back at least as early as 1922, when Chao obtained pitch contours of lexical tones of Beijing Mandarin (and many other Chinese dialects/languages) that are rather close to what we know today, using very rudimentary instrumentation [4]. Today, however, despite rapid technological advances that have made instrumental acquisition of F_0 and many other acoustic measurements of speech fairly easy, progress in the scientific understanding of tone remains very uneven in the field. The main obstacles to more comprehensive progress in tone research is therefore not in the lack of technological means, but in the mindset of our discipline. This paper explores ways to improve tone research by reviewing general methodological principles, specific biological mechanisms related to tone, and the newly proposed articulatory-functional principle.

2. Methodological principles

The goal of science is to understand how things work. This can be achieved only by uncovering the underlying mechanisms in terms of causal relations. This is true no matter what approach is taken toward a subject matter. For tone research, it should be true whether one wants to understand the tonal phenomena from a phonological or phonetic perspective. The establishment of causal relations requires meticulously thorough investigations, and methodology is often vital to the success of such pursuit. The following sections discuss four methodological principles that are critical for speech research in general and tone research in particular: *experimental control*, *spare no details*, *always look for actual mechanisms* and *one step at a time*.

2.1. Experimental control

Experimental control means to make observations by manipulating the factor being investigated while keeping others constant. Observed variations can then be directly attributed to the manipulated factor. This simple idea is one of the principal foundations of modern science. Its importance can be seen in the fact that if the research concerns a matter of life or death, such as in medicine, extremely stringent requirements regarding experimental control are imposed.

Unfortunately, experimental control has not always been rigorously applied to linguistic research in general, and tone research in particular. This may be excusable for early field work due to methodological limitations. The situation is not much improved in recent years, however, except in the cases of a limited number of languages. A large amount of the phonological analyses of tone are still based on data obtained with less than rigid experimental control. Furthermore, there has been a rise in popularity for examining spontaneous speech as a way of studying tone and intonation. The trend is driven in part by the belief that "laboratory speech" is unnatural, too careful, lack of variability, etc., and so can tell us little about how real people speak in their everyday life. Interestingly, a similarly motivated trend has been seen in memory research in favor of directly examining memory phenomena in naturalistic situations [1]. But as pointed out by Banaji and Crowder [1:1189]:

> … the multiplicity of uncontrolled factors in naturalistic contexts actually prohibits generalizability to other situations with different parameters. The implication that tests in the real world permit greater generalizability is false once the immense variability from one real-world situation to another is recognized.

If the ultimate goal of tone research is to improve our understanding of speech, then true insight can come about only when the causal relations behind the tonal phenomena are uncovered. A causal relations can be established only when it is reasonably certain that all the other factors that might also have been involved remain constant. In spontaneous speech, many factors are involved at the same time, but few can be known for certain. Thus the richness may actually form impenetrable obstacles to true understanding.

Based on the principle of experimental control, the most effective observation may be made in minimal contrast comparisons. For example, when all the other factors are kept constant while the surrounding tonal contexts are systematically varied, and when the F_0 contours are overlaid in the same plot, as shown in Fig. 1, not only is the nature of the contextual tonal variations but also that of the tone itself is revealed [64], as will be discussed in 3.3.

Suppose that the syllables in Fig. 1 differed not only in tone, but also in initial consonant, rhyme, position in word and/or sentence, and also in the carrier sentence in terms of utterance modality (statement vs. question), focus, position in a paragraph or conversational turn, and in the emotional and attitudinal status of the speaker. How, then, can one be sure which part of the observed F_0 pattern is due to the tonal differences and which is due to any of the other factors?[1]

It is not even true that speech samples obtained with experimental control is any less "natural" than spontaneous

[1] To make things worse, it is not uncommon to find in the literature theoretical arguments based on isolated sentences whose source and environment are almost totally unspecified: whether it is from a dialogue or a monologue, its position in a paragraph or conversational turn, the preceding and following sentences, the recording conditions, and sometimes even the gender of the speaker. One cannot help but wonder what is the validity of the arguments based on this kind of observation.

speech. First, even the most "monotonous" laboratory speech is still human speech produced by real speakers. In fact, some of the best insights have been obtained from such laboratory speech [12,43,52]. Secondly, what might be missing in certain experimental conditions are factors that are not automatically elicited. But that does not mean that they can never be controlled. In fact, virtually any factor is potentially manipulatable by experimental means, limited perhaps only by ethnical concerns in very limited situations (e.g., those linked to extreme emotions). For Mandarin, for example, not only has contextual tonal variations been experimentally examined, but also tonal variations due to several nontonal factors have been investigated through experimental manipulations, including consonantal perturbation [62], vowel intrinsic F_0 [50], syllable structure [22,65], focus [6,66], sentence modality [36], speech rate [29,67], topic [55] and syllable grouping [74], etc.

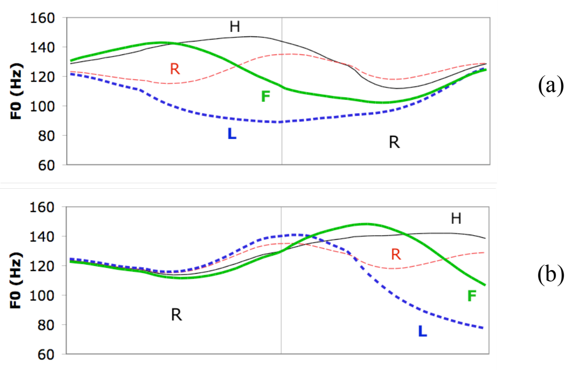

Fig. 1. Time-normalized mean F_0 of Mandarin R tone preceded by four different tones (a) or followed by different tones (b). Each curve is an average of 48 repetitions by 8 male speakers. Data from [64].

This is of course not to say that spontaneous speech has no value in tone research. It is useful in motivating new hypotheses, raising questions about existing theories, and testing the validity of proposed models, especially the quantitative ones.

2.2. Spare no details

Experimental control itself does not fully guarantee sure progress in our knowledge. Just as important, scientific understanding can be only as accurate as the level of detail we choose to use in making observations. In the case of tone research, if we take measurements from only a limited number of points, such as at the F_0 peak, valley, the center of a vowel, etc., from the speech samples we study, although certain gross patterns can be observed, the causal relations among the contributing factors may still remain vague. In Fig. 2a and 2b, for example, only 1 or 2 measurement points are taken from each tone-bearing syllable in Mandarin 5-syllable sentences. Although the experimental design has guaranteed that all the extraneous factors are tightly controlled, still not much can be learned in terms of the exact nature of the contextual tonal variations. The picture becomes much clearer when more observation points are obtained, as seen in Fig. 2c, where 17 F_0 points are displayed for each syllable.[2] As will be discussed in 3.3, it is based on this kind of detailed observations that the underlying mechanisms of tone become better understood.

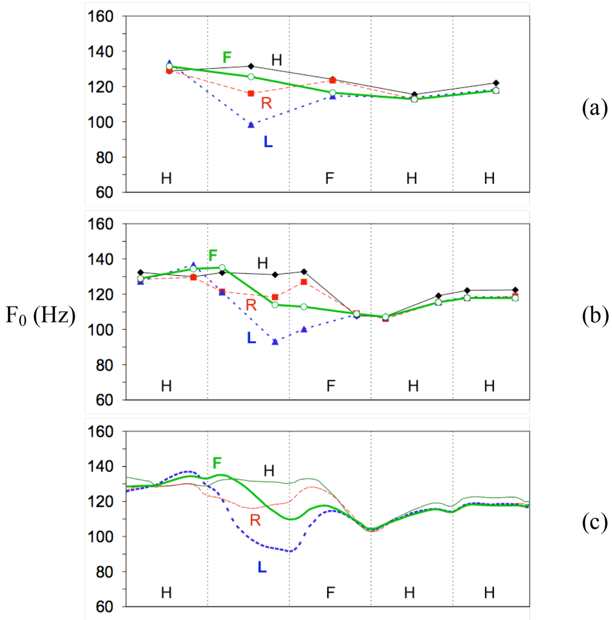

Fig. 2. Plots of mean F_0 values taken from 5-syllable sentences in Mandarin. (a), (b) 1 or 2 measurement point(s) per syllable, joined by interpolation lines. (c) 17 measurements per syllable. Data from [66].

So, as long as feasibility allows, which itself should be a constant goal for improvement, observations and measurements should be as fine-grained as possible. Thanks to the availability of computer programs like Praat, extracting detailed F_0 contours is no longer something only a few sophisticated laboratories can achieve. Anyone with a laptop and a microphone can perform detailed F_0 analysis even in a field trip. See [60] for a Praat script that allows one to obtain, among other things, time-normalized F_0 contours like the ones shown in Fig. 1 and 2, and to take various measurements for systematic analysis.

2.3. Always look for actual mechanisms

To understand the observed phenomena in terms of causal relations, it is not enough to simply describe the phenomena we observe. It is more important to reveal the mechanisms behind the phenomena so that we can make generalizations and predictions. Whether we can uncover the actual mechanisms depends much on whether we are consciously and persistently looking for them. In this regard, we should be especially vigilant against the tendency to resort to reification as a shortcut to true understanding. Reification refers to the act of treating an abstraction as if it were a real, concrete thing [41]. In research, it is often tempting to give an observed phenomenon a descriptive name and then treat it as an actual entity. As pointed out by Ohala, giving in to such temptation may actually hamper real progress [41:161]:

> The trouble with reification, then, is that it puts the problem solvers one step further away from the real causes of the problem; their limited resources – both psychic and material – may be wasted on the nonexistent reified entity rather than the true root causes of the problems they face.

A case in point is the phenomenon of F_0 downtrend. It has

[2] 17 points per tone is not a magic number. Fewer points may be enough for each tone, as long as the continuous F_0 movements are clearly visible.

been given names such as declination [7] and downdrift [17]. Beside being used just as convenient descriptive terms, these names have been reified into actual mechanisms in both tone [51] and intonation research [8]. As demonstrated in [66], however, there are at least three independent mechanisms behind such downtrend: downstep, non-final focus and topic. Of these, downstep further consists of two separate mechanisms: anticipatory raising [14,32,64] and carryover lowering. More recent data further show that utterance modality also contributes to such downtrend [36].

Another case is the phenomenon known as tone spreading [23], which has been reified into a phonological rule [15]. As has been pointed out, the name is at best a cover term for two very different processes: a mechanical carryover effect and a likely reassignment of the underlying pitch target [70]. The second process is still quite unclear, however, as the target reassignment account in some cases is supported so far only by anecdotal evidence [70]. Investigations with proper experimental controls are needed.

As will be discussed later, many other widely accepted notions in tone related research may have also been reified entities rather than actual mechanisms or operable units. These include such well-known terms as hierarchical prosodic structure, pitch accent, prominence, rhythm, feature, etc.

In general, real mechanisms operate on specific physical, physiological, neural or informational principles. Symbolic operations, which are still widely used in tone research, are unlikely representative of real mechanisms. There are assumptions that symbolic operations represent *cognitive* processes [46]. But cognitive processes are by definition *neural*. As such they cannot be symbolic at the actual operational level. Symbolic operations such as tone spreading, association line, feature geometry, etc., therefore cannot not provide insight into the real mechanisms behind various tonal phenomena.

Closely related to the problem of reification is that of circularity. Known in logic also as *begging the question*, circular reasoning takes what is to be explained as part of the premise. For example, as pointed out by [41], the theory of markedness, which is also widespread in tone research, is circular because it uses the commonness of sounds or feature combinations as an explanation for why certain sounds or features are more common than others. As has been argued, some of the most important markeness constraints arise from the basic mechanisms of motor control in speech production, and thus do not need to be assumed as part of the innate knowledge as assumed by the markedness theory [71].

2.4. One step at a time

As seen in the case of F_0 downtrend, a tone related phenomenon may have multiple mechanistic sources. Because in each controlled experiment only a limited number of factors can be manipulated, it is often impossible to reveal all the involved mechanisms at once. In fact, such situations are the rule rather than the exception. An effective strategy is to first isolate the most tangible ones, and then proceed to the more difficult ones. The effectiveness of such strategy lies in the fact that every time a particular mechanism is identified, its likelihood of acting as a hidden confounding factor is reduced. For example, among the tone related factors, tones themselves are the easiest to control, as they can be elicited simply by using different words. It is also relatively easy to control focus, as it can be elicited by different leading questions [66]. The focus effect on F_0 can then be seen as the variations that cannot be attributed to the tonal variations, as can be seen in Fig. 3a and will be explained in more detail in 4.4.

It is critical, however, that once a mechanism is recognized, variations clearly due to it not to be re-attributed to other mechanisms. As can be seen in Fig. 3b (as well as in Fig. 1-2), the tonal differences necessarily lead to differences in the alignment of the F_0 peaks relative to the syllable. It would then be unnecessary to posit additional mechanisms that control the alignment of the F_0 peaks and valleys [68]. Also based on this understanding, we should be weary of any all-encompassing theory that claims to account for everything without specifying any non-symbolic mechanisms.

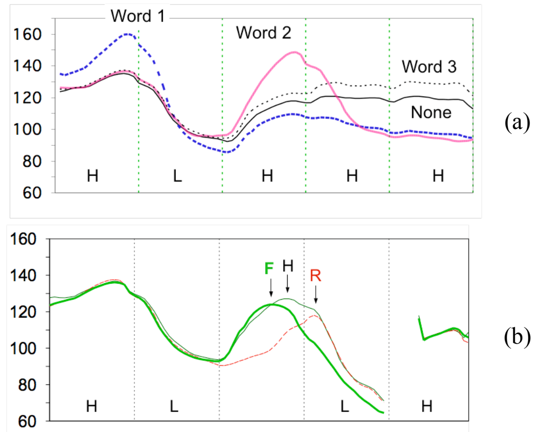

Fig. 3. Time-normalized mean F_0 contours of five-tone sequences. (a) HLHHH with focus on word 1, 2 (mono-syllabic), or 3, or no focu. (b) HLxLH where x is F, H or R. Vertical arrows point to F_0 peaks. All curves averages of 24 repetitions by 4 male speakers. Data from [66].

3. Biophysical mechanisms

Although there is much literature on the basic physiology of pitch production [19,20,40,53,79], tone research has not taken the physiological properties of the pitch production process seriously until recently. One reason is that many tonal phenomena are not readily explained by the general knowledge about the physiology of pitch production. More specific mechanisms need to be taken into consideration. The following discussion will take an overview of some of known issues. In each case, biophysical interpretation is attempted, but more in-depth research is clearly needed.

3.1. Maximum speed of pitch change

In a tone language in which every syllable carries a tone, underlying pitch patterns have to shift 5-8 times per second [66,10]. Existing physiology literature, however, does not directly tell us how demanding such frequent pitch shift is at the muscular level [19,53,79]. Evidence obtained at the behavior level [72] indicates that the minimal time needed depends on the magnitude of the pitch excursion, as approximated by the following linear equations:

$$t = 89.6 + 8.7\, d \qquad (1)$$

$$t = 100.4 + 5.8\, d \qquad (2)$$

where t is the amount of time it takes to complete the pitch shift, and d is the size of pitch shift in semitone. From (1) and (2), it takes about 100 ms for an average speaker to complete even the smallest amount of pitch change; and raising or lowering pitch by 4-st takes at least 124 ms. The finding has at least the following implications.

1. At the speed of 5-8 tones per second, each tone is given

only 125-200 ms. Much of the F_0 in speech would therefore be mostly transitions from one tone to next.
2. The maximum speed of pitch change is often approached. The fastest pitch change reported by [3] is comparable to the maximum speed of pitch change at similar pitch shift intervals reported by [72]. The maximum speed of pitch change reported by [72] also matches the speed of pitch change in the dynamic tones (R, F) in Mandarin [66].
3. Even so, in many cases, the speed of pitch change is still not fast enough. As a result, certain tones can be totally flattened [29,63].
4. In many languages, this has led to prohibition of contour tones in certain situations [16,80] because they routinely require two F_0 movements within one syllable [69].

3.2. Synchronization of laryngeal and supralaryngeal movements

The control of F_0 and the spectral aspect of speech involves separate mechanisms and they encode different information in speech. This however does not mean that the temporal controls of the two are independent of each other, as assumed in autosegmental phonology [15]. Evidence for their temporal coordination comes from several sources.

1. To carry out two concurrent cyclic motor movements, performers have few choices in terms of inter-movement phase relations [27,37]. At low speed, the phase angle between the two movements has to be either 180º, i.e., starting one after the other is half way through, or 0º, i.e., full synchrony between the two. At high speed, however, only the 0º phase angle is possible. As just mentioned, pitch movements in speech are frequently as fast as possible. Laryngeal and supralaryngeal movements therefore tend to be fully synchronized.
2. Even if consecutive syllables are not cyclic in nature [54], there is a strong tendency for two non-cyclic motor movements to be fully synchronized [28].
3. Tone production in Mandarin has been found to be synchronized with the syllable: the movement toward each tonal target starts at the syllable onset and ends at the syllable offset, as can be seen in Fig. 1-3.
4. Such synchronization is maintained regardless of the voicing of the initial consonants [62] or whether the syllable has a nasal coda [65].
5. The tone-syllable synchronization may reflect an even more fundamental mechanism of temporal organization in speech. According to the recently proposed *time structure model* of the syllable [71], the onset of the articulatory movement toward a tonal target, as well as those toward the first consonantal and vocalic targets of a syllable, all start about 50 ms ahead of the conventional syllable boundary, e.g., closure onset in a stop, affricate, nasal or lateral, or frication onset in a fricative.

3.3. Target approximation

The Target Approximant (TA) model [75], as is illustrated in Fig. 4, was proposed based on (1) detailed F_0 contours obtained through controlled experiments [64-67], (2) the understanding that changing pitch takes time [72], and (3) that there is a strong tendency toward synchronization between concurrent movements [27,28,37].

According to the TA model, each tone is associated with a pitch target in the form of a simple ideal pitch pattern. In most cases the target is a linear line which is either level (static) or sloping (dynamic). More complex target configurations are also possible [75]. The process of tone production is then one of asymptotically approaching the pitch target, starting from the initial F_0 at the onset of the syllable. The approximation terminates at the offset of the syllable, whether or not the target has been reached.

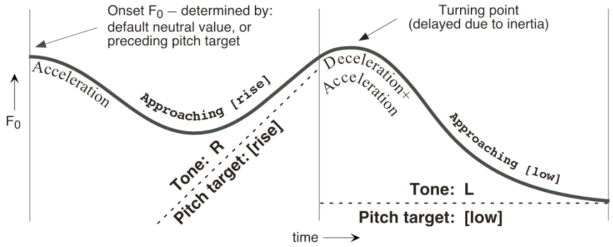

Fig. 4. Illustration of the Target Approximation (TA) model. The vertical lines represent syllable boundaries. The dashed lines represent underlying pitch targets. The thick curve represents the F_0 contour that results from asymptotic approximation of the pitch targets.

The TA model has been used to account for patterns of contextual tonal variations in Mandarin as well as in some other languages [58]. A quantitative version of the TA model has also been developed and the initial testing has yielded encouraging results [47].

3.4. Total pitch range

According to [79], a speaker's conversational pitch range spans about 2 octaves (24 st). Data from a more recent study show that the mean non-singing pitch range of American English speakers, even when not including the falsetto register, is about 3 Octaves (35.4 st) [21][3]. In Mandarin, F_0 variations due to the four Mandarin tones span only about one octave at any particular sentence position [66]. Also, the tonal pitch range is in the lower portion of the total pitch range. In [66], the highest mean pitch in sentences with no focus is about 140 Hz for males and 300 Hz for females. With focus it is about 164 and 342 Hz for males and females, respectively. These are much lower than the highest mean F_0 in [21]: 340 Hz for males and 500 Hz for females.

The lower pitch limit, however, is frequently reached during normal speech. For reasons not yet fully clear, F_0 around the lower pitch limit is often associated with creaky voice. In Mandarin, for example, the L tone is known to be often produced with creaky voice. There is thus a question as to whether such creakiness is part of the primary property of the L tone or is just a byproduct of very low pitch. It has recently been demonstrated that Vietnamese tones are actually realized as three-way voice quality contrasts: modal, breathy and creaky, whereas F_0 patterns corresponding to the tones are much more variable [44]. In Mandarin, it is the low F_0 in the L tone that is highly consistent, while voice quality actually varies between speakers from modal to creaky to breathy and even complete break of phonation [81]. Thus reaching the lower limit seems to be the cause of the voice quality changes rather the latter being the primary property of the L tone.

3.5. Post-low bouncing

Related to the lower limit of speakers' pitch range is the newly found phenomenon of post-low bouncing. That is, following a

[3] Courtesy of Douglas N. Honorof for the numerical data.

very low pitch, F_0 tends to rise to a higher level than usual before falling back to normal. The phenomenon has been reported for both Mandarin [7] and English [45]. In each case, there is a relatively constant time interval between an F_0 valley and the following peak. In English the interval is about 191–202 ms [45]. In Mandarin, it is about 175 ms when the post-L tone is H, and 232 ms when it is Neutral. There have to be two Neutral tones following the L for the interval to be measurable, of course, because the duration of a Neutral tone syllable is typically much shorter than 200 ms [7].

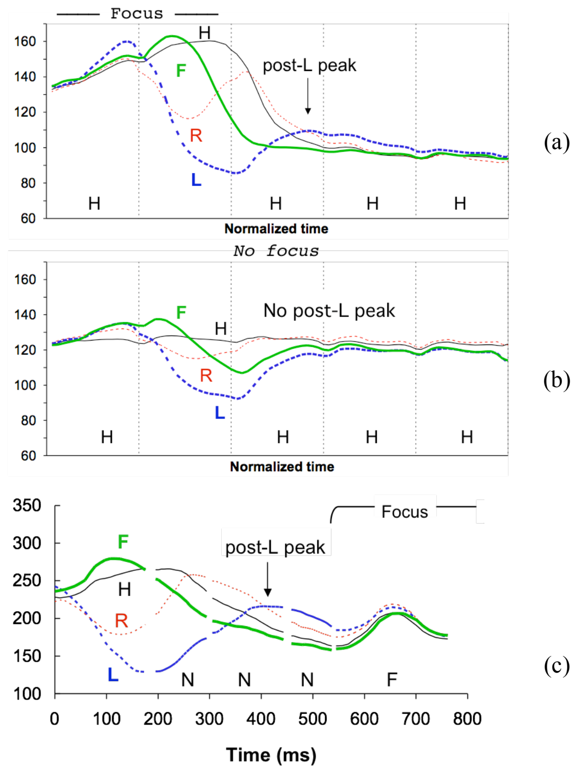

Fig. 5. Mandarin F tone following four different tones. Top: no narrow focus in the sentence; Bottom, focus on the F-carrying syllable. Data from [7].

The underlying mechanism of the post-low bouncing is unclear. But there are some hints. First, it seems to vary with two interactive factors. The first is the height of the low F_0. The lower it is, the more likely the bouncing occurs. The second is the nature of the following tone. The weaker the tone, the more likely the bouncing occurs and the longer it lasts. In Fig. 5a the bouncing can be clearly seen in the H tone when the preceding L is focused and thus becomes even lower in F_0. But the bouncing cannot be seen in Fig. 5b where the L is not focused. In Fig. 5c, the bouncing occurs even though the preceding L is not focused, presumably because the Neutral tone is weak, as will be discussed in 4.3.1.

It is known that very low F_0 is produced with the contraction of the extrinsic laryngeal muscles such as the sternohyoids, sternothyroids and omohyoids [40]. It is possible that the sudden disengagement of these powerful muscles may remove the force antagonistic to the cricothyroid muscle, whose continued contraction temporarily overcorrects the F_0. It has also been found that the contraction of the extrinsic laryngeal muscles lowers F_0 by sliding down the cricoid cartilage over a backward bent in the cervical spine, forcing it to rotate forward, hence shortening and relaxing the vocal folds [20]. Terminating the contraction of the extrinsic laryngeal muscles may result in the cricoid cartilage rotating backward when sliding up over the bent in the cervical spine, thus over-lengthening the vocal folds. More research is needed to reveal the mechanics of post-low bouncing.

3.6. Intrinsic F_0 of vowels and F_0 perturbation by consonants

Other things being equal, different vowels are known to have different F_0 values [57]. Such variation is mostly related to vowel height: the higher the vowel, the higher the F_0. The intrinsic F_0 difference is in the range of 1.65 semitones or 15 Hz [57]. Intrinsic F_0 has been found in both tone and non-tone languages. Being small in magnitude, it does not usually affect either the local or global F_0 patterns in a significant way. There is also evidence that its magnitude becomes smaller in connected speech [31]. Nevertheless, when analyzing an effect on F_0 by other factors, intrinsic F_0 also needs to be taken into consideration.

A non-sonorant consonant perturbs F_0 in two ways. First, it interrupts the otherwise continuous F_0 movement. Second, it raises or lowers the F_0 of the adjacent vowels. While the F_0 raising and lowering have been widely known, the effect of F_0 interruption is often not carefully considered. It has often been taken for granted that because of the voicing interruption, consonants are not part of the tone-bearing unit. There is now evidence, however, that the voicing interruption does not really disrupt the TA process discussed in 3.3. That is, the syllable-synchronized sequential target approximation operates in the same way whether or not voicing continues through the consonant [62]. Thus the effect of voiceless consonants is only to bring rather local perturbations without changing the basic mechanism of tone production [62,73].

Consonants may also affect the height of the F_0 in both the preceding and following vowels. While the latter effect is well known [18,34], consonantal effect on the F_0 of the preceding vowels is not yet well documented. But a recent observation has revealed certain local effects in English [73]. Further research on this effect is needed.

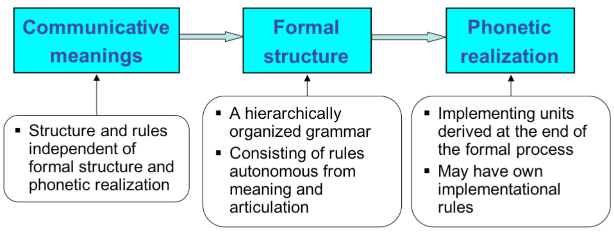

Fig. 6. A schematic of conventional conceptualization of the speech production process.

4. The articulatory-functional principle

The principle of *always look for actual mechanisms* is applicable in dealing not only with lower level processes like those just discussed, but also with system-wide processes. For a long time, the dominant view about the general process of speech production is that it consists of three separate steps, as depicted in Fig. 6. In this view, the communicative meanings (left) have their own structure and rules independent of other processes. Their link to the surface acoustic form is at best rather obscure. The transmission of communicative meanings is then done through a formal process (middle) that is largely autonomous from both meanings and articulation. It is after the completion of this formal process that the output of the phonological derivation is implemented by the phonetic process (right), which also has its own rules. Such a view is most clearly expressed for prosody and intonation in [2,45] and for tone in [15]. Based on this view, the task of speech production is first and foremost to guarantee that a set of formal requirements are met (viz. that the output is grammatical). And understanding speech is done not by directly accessing the meaningful components, but by first parsing a phonological structure [2].

The problem with such conventional view is that, while the involvement of communicative meanings and articulatory mechanisms in speech is inescapable as most people would agree, the involvement of a formal entity with largely autonomous internal structure and rules actually requires strong justifications. Following the principle of *one step at a time*, one should always start with the most inescapable, in this case the communicative functions and articulatory mechanisms, and see how many observed patterns can be accounted for. If, after doing so, much is still left unexplained, it may then be justifiable to consider additional possibilities.

What, then, are the communicative functions and articulatory mechanisms? Again, it is helpful to first identify the most tangible ones. For the communicative functions, as will be discussed in subsequent sections, what is already known are lexical contrast, focus, new topic, sentence modality, and chunking/grouping. For the articulatory mechanisms, most have been discussed earlier, but a critical one — target approximation, needs to be further discussed, as it provides the likely core mechanism for encoding the communicative functions.

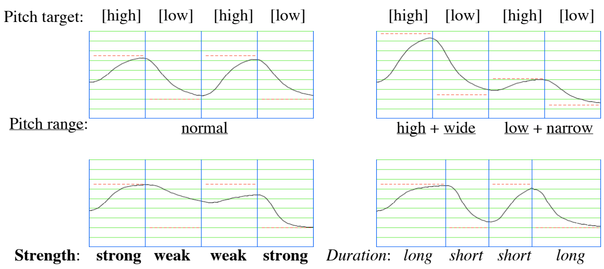

Fig. 7. Illustration of the effects of pitch targets, pitch range, strength and duration. Generated by the Java applet accessible at [61].

4.1. Target approximation as an encoding mechanism

The target approximation process represented by the TA model provides not only a mechanism of tone production, but also a mechanism for encoding different types of information. That is, various parameters of the TA process can be manipulated, and the manipulations can generate identifiable changes to surface F_0 contours. At least four basic parameters can be recognized: pitch target, articulatory strength, pitch range and duration. Pitch target specifies the properties of the local pitch goal associated with each syllable (or mora). Articulatory strength specifies how quickly a pitch target is approached. Pitch range specifies the pitch span within which a target is approached. And duration specifies the time interval assigned to the approximation of a target. The effects of manipulating these parameters can be seen in Fig. 7, which shows F_0 contours generated by an earlier quantitative version of the TA model [61].

4.2. The PENTA model

PENTA is the acronym for *Parallel Encoding and Target Approximation*. A schematic sketch of the model is shown in Fig. 8. The key idea of the model is that F_0 conveys multiple layers of communicative functions that are parallel to each other (far left). Each function is directly coded by an encoding scheme (second left) which specifies the values of the articulatory parameters (middle) for the TA process (second right). The TA model then generates surface F_0 through the process of *syllable-synchronized sequential target approximation* [70]. Thus surface F_0 is not directly specified by any phonological or phonetic units. Rather, it is generated by the TA process whose control parameters are specified by the encoding schemes, each corresponding to a specific communicative function. There is therefore no place for a self-contained hierarchical structure with internal rules that govern the generation of well-formed surface output.

Some of the communicative functions and their encoding schemes have been investigated in recent research. But many remain unclear. The following discussion will address each of them briefly, pointing out in particular where further research is needed.

4.3. Lexical contrast

Lexical tones in a tone language serve a clear communicative function: to distinguish words from one another [5,78]. In this sense, tones are not really different from consonants and vowels. Also like consonants and vowels, tones are encoded by controlling the most local aspect of articulation in the time domain. However, as discussed earlier, because it takes at least about 100 ms to change pitch by any significant amount, tonal pitch control is typically done at the level of the syllable [71]. Although the mora is also a potential level of pitch control for languages like Japanese [78], it is possible, pending future research, that even that is implemented at the level of the syllable.

In the TA model, the most local component of pitch

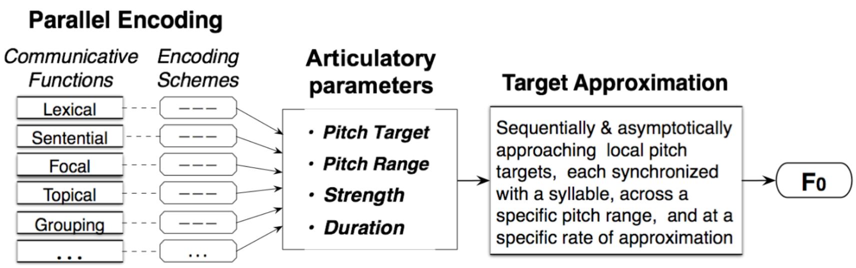

Fig. 8. A sketch of the PENTA model. See text for explanations. The unnamed block at the bottom left indicates communicative functions yet to be identified. Adapted from [70].

control is the pitch target, which is sufficiently defined in terms of height and slope in the cases investigated so far [70]. When the slope is zero, the target is static; otherwise it is dynamic. For example, the four Mandarin full tones in non-final positions appear to have the pitch targets [high], [low], [rise] and [fall], in which the latter two are dynamic [75]. Similar target analysis has recently been done for Cantonese [58-59], for which very similar target approximation patterns have been found.

4.3.1. Neutral tone

Tonal distinction can be encoded not only by the height and slope of the pitch target, but also by articulatory strength which determines the speed at which the target is approached (Fig. 7). Because full articulatory strength is likely often used due to the maximum speed of pitch change, the only major strength manipulation would be a substantial reduction. Other things being equal, a reduced strength slows down the target approximation. A likely case is the Neutral tone in Mandarin. A syllable carrying the Neutral tone is traditionally considered to be toneless, because its F_0 appears to vary depending on the tone of the preceding syllable. As can be seen in Fig. 5c, although the F_0 of the first Neutral tone is much influenced by the preceding tone, before the end of the syllable, the F_0 contours have all turned toward a common mid level value. In the two subsequent Neutral tone syllables, the F_0 contours continue toward this value. The only exception is the one after L due to the post-low bouncing discussed in 3.5. But even in that case, when the effect is over, F_0 also turns toward the common value. Nevertheless, even by the end of the third Neutral tone, substantial differences related to the full tone in the first syllable still remain.

These patterns suggest that, first, the Neutral tone likely has its own pitch target. Otherwise, there would not have been such a clear tendency for the F_0 contours to converge to a common value starting even from the first Neutral tone. It has been calculated that this common value is at the mid level of the position-specific tonal pitch range [7]. Secondly, the approximation of the Neutral tone target is done with a much reduced effort as compared to a full tone. Otherwise, the influence of the first tone would not have faded so slowly, remaining visible even by the end of the third Neutral tone.

It has also been shown that even in English, a non-tone language, the F_0 of an unstressed syllable is best explained as resulting from approaching a [mid] target with weak articulatory force [76]. It is thus possible that similar encoding strategy is employed by many other languages, tonal or non-tonal.

4.3.2. Target reassignment versus parallel encoding

Despite the lexical contrast function they serve, a lexical tone sometimes behaves as if its pitch target is simply changed in certain situations. A clear example is the L-tone sandhi phenomenon in Mandarin, by which the first L in a LL sequence is changes into a form perceptually nondistinct from the R tone [42,56]. Detailed acoustic analysis show that the F_0 contour approaches a [rise] target from the syllable onset, just as in an underlying R [64]. Thus there seems to be a genuine *reassignment* of the pitch target, although the exact reason for such reassignment is still unclear.[4]

This kind of tonal variation is different from other cases of tonal variability. In addition to the contextual variations which can be accounted for by the TA model, there are also many variations related to other communicative functions. In light of the PENTA model, different communicative functions are parallel to each other, although there may be interactions among them. Changes in F_0 due to the non-lexical functions therefore should not be treated as tonal changes as such, unless there is clear evidence that the pitch target is reassigned. As will be discussed in the next few sections, at least in Mandarin, F_0 variations introduced by focus and sentence modality do not lead to reassignment of pitch targets. From the functional perspective, therefore, it is inappropriate to refer to F_0 variations due to those non-lexical functions as tonal [15].

4.4. Focal contrast

Focal contrast as an important communicative function involves robust F_0 manipulation for its encoding [66,70]. In a tone language, such manipulation introduces extensive variations to the F_0 contours generated by lexical tones, as illustrated in Fig. 3a. Knowledge about focus is thus vital to understanding tones produced in continuous speech. The following are a number of key issues related to focus.

1. The encoding scheme of focus is global rather than local. It consists of three regions of pitch range specifications: expansion in focused region, suppression in post-focus region, and neutral in pre-focus region [77]. Measuring and interpreting F_0 for tones should therefore be concerned about not only the effects on the focused word, but also those on the post-focus words.
2. Focus is to highlight new information or to deemphasize wrong information. It is thus easily elicited by information contained in either the current or preceding sentence. Under experimental conditions, words in a constant carrier tend to be focused, because they are the only ones changing from trial to trial. By the same token, when reading a list of sentences, there is a natural tendency to focus on the portion of a sentence that differs from the preceding sentence [66]. Also, certain sentence formations, such as wh-question or those containing words like “only”, “not”, etc., are effective focus inducers [24].
3. Final focus is often not effectively encoded, and its influence on F_0 tends to be smaller than earlier focus [36,66,77]. This is presumably due to a conflict with the utterance modality function [36], as will be discuss in 4.6.
4. The notion of broad focus is problematic. It is used to refer to the focus state of sentences with no narrow focus [30], for which it is said to have a default final focus. But final focus is both acoustically and perceptually different from no focus [36,38,48,66,77]. Also it is possible to contrast real broad focus that emphasizes the whole utterance with final focus [26]. It is true that final focus is perceptually less distinct than earlier focus [36], but that is probably what has made some researchers believe that focus occurs in every sentence. If forced to identify a focus in a sentence with no emphasized word, chances are one would point to the sentence final word whose focus status is most easily confused with that of a real final focus.

4.5. New topic

New topic is a function that contrasts a newly introduced topic

---

[4] There is evidence that this is a historical change that started centuries ago and slowly spread across many northern dialects [25]. This makes the true mechanism of the change hard to unearth, as it is difficult to know the exact pitch value of the L tone at that time.

with an old one [70]. It is closely related to turn taking in conversation [39]. It raises (but not expands) the pitch range of the initial portion of an utterance. A recent study demonstrates that new topic differs from initial focus in that (a) it does not lower the floor of the initial pitch range, and (b) it does not suppress the pitch range of the subsequent words [55]. The exact function of new topic is still not quite clear, but it may be related to drawing the attention of the listener. As such it may involve not only raised pitch range, but also increased intensity [70]. In any case, the pitch range raising by new topic certainly affects F_0 measurement for tones. In fact it likely contributes to the phenomena of downdrift, declination and probably even downstep [70].

4.6. Utterance modality

Also known as sentence type, utterance modality contrasts interrogative meaning from declarative meaning, or, in other words, question from statement. There have been much research on the acoustic cues for utterance modality. A consistent finding is that F_0 is raised toward the end of an utterance [36]. Although full consensus has not been reached on a number of issues, for tone research, some of the known effects of utterance modality cannot be ignored [36,70].

1. In contrast to new topic, utterance modality is encoded by manipulating pitch range toward the end of an utterance. Recent evidence shows that the contrast in F_0 between interrogation and declaration accelerates over time and reaches a maximum at the end of the utterance [36]. Thus the effect on tone is not just at the final location, but also earlier in an utterance, although the magnitude is smaller toward the start of the utterance, other things being equal.
2. The modality contrast affects not only the pitch range of questions, but also that of statements. In the latter case, it contributes to various phenomena of downtrend as discussed earlier. Also the final acceleration is closely related to the phenomenon of final lowering [35].
3. The modality contrast is enhanced by focus such that the difference between question and statement is increased from the focused word onward [36]. Thus tone investigation involving utterance modality needs to make sure that focus is properly controlled.
4. Some languages may employ additional encoding strategies to enhance the modality contrast. In English, for example, in addition to the focus-conditioned pitch raising, the stressed syllables may be reassigned different pitch targets [11]. It is possible that some tone languages may also employ additional encoding strategies for the modality contrast.
5. It is well established that many sentences with question syntax are said without a rising pitch pattern [49]. This indicates that the modality contrast is not syntactic in nature. Thus the real communicative meanings being discriminated need to be further researched.

4.7. Duration related functions

Durational variations of tone are related to several communicative functions. Focus has been found to lengthen the focused words without affecting the duration of either the pre-focus or post-focus components [6,66]. In Mandarin, the neutral tone shortens the duration of a syllable to about 61% of the regular syllable [7]. Many Chinese languages have a tone class known as the entering tones [78]. They are characterized not only by pitch patterns, but also by robustly shortened duration. There are also suggestions that they are actually shorter versions of the corresponding non-entering tones, their F_0 variations being entirely attributable to the shortened duration [33]. Durational contrasts also exist in many other tone languages. Whether directly related to tone, they certainly have various effects on the acoustic realization of tones.

Another duration related function, whose exact nature still remains unclear, has to do with grouping syllables into chunks in an utterance. This function is related to phenomena such as rhythm, foot structure and prosodic hierarchy. Recent evidence shows that there exists a within-foot duration pattern in Mandarin such that syllables at the edges of a foot are lengthened. It is further found that these duration patterns can account for the F_0 variations that, on the surface, appear to be due to changes in articulatory strength [74]. Thus syllable grouping may have little to do with strength manipulation. More importantly, such duration manipulation for the sake of grouping syllables is apparently parallel to encoding schemes for other functions, and is unlikely to form a prosodic hierarchy for governing all tonal and intonational components [15,30].

4.8. Other functions and perception related phenomena

The above discussion certainly does not cover all the tone related functions. Additional ones would include at least various attitudinal and emotional functions, which we still know very little about. Also perceptual issues have not been separately addressed due to lack of space. Nonetheless, the methodological principles, biomechanical mechanisms and the articulatory-functional principle should be just as applicable when investigating both the additional functions and perception-specific phenomena.

5. Putting it together

Now that we have gone through the major articulatory mechanisms and communicative functions that have been investigated, we are at a point where we can put these things together and see how many of the observed patterns can be accounted for. Fig. 9 displays mean F_0 contours of Mandarin five-syllable utterances with the tone sequences of HHHHH (dashed curve) and HLHHH (solid curves), and the following observations can be made.

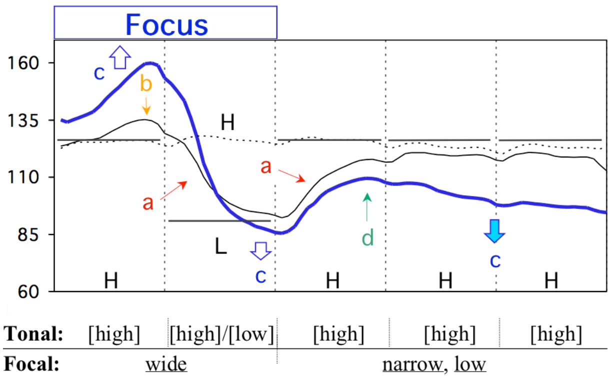

Fig. 9. Time-normalized mean F_0 contours as illustration of the encoding schemes of tone, focus realized with various articulatory mechanisms, The short horizontal bars illustrate the pitch targets of the tones. The block arrows point to the pitch range manipulations by focus. The thin arrows point to various mechanical effects.

1. When all the syllables in the utterances have the H tone and when there is no focus (dashed curve), F_0 contour is

mostly flat, because the pitch targets are all [high].

2. A slight overall downward tilt can nevertheless be seen, which is likely due to both new topic (for the utterances were said in isolation) and the declarative modality, as discussed in 4.5 and 4.6.
3. When the tone of the second syllable changes into L, its pitch target becomes [low] (short horizontal bar). F_0 in the second syllable approaches this target asymptotically, reaching the lowest point around the end of the syllable (Arrow a). The asymptotic approximation can be also seen in syllable 3, where [high] is approached after the low F_0 due to the preceding L. These patterns can be fully accounted for by the TA model (3.3).
4. The production of the [low] target also causes anticipatory raising [14,32,64], as can be seen in the slightly higher F_0 of the thin solid curve than the dashed curve in syllable 1 (Arrow b). While its mechanism is still unclear, the phenomenon can be clearly separated from other phenomena [64,66].
5. In addition to raising the F_0 of the preceding H, the [low] target of syllable 2 also lowers the F_0 of syllable 3. This lowering is never fully recovered even by the end the utterance. The exact nature of this extensive lowering is yet unclear.
6. When the first two syllables are focused, their pitch range is expanded. The expansion also increases the size of the rising ramp in syllable 1 and the falling ramp in syllable 2 due to the increased distance to be covered from the syllable-initial F_0 (see equations 1, 2). Meanwhile, the pitch range of subsequent syllables are lowered due to post-focus pitch range suppression (Arrow c).
7. The very low F_0 due to expanded pitch range in the L tone in syllable 2 also leads to the post-low bouncing (3.5) seen most clearly in syllable 3, which fades away gradually in syllables 2 and 3 (Arrow d).

What the above example and the earlier discussion have demonstrated is the explanatory power of the articulatory-functional principle in conjunction with the individual underlying mechanisms. They also show that F_0 contours in Mandarin, and probably in other languages as well, are unlikely to have been generated by either a strictly linear process [45], or a strictly superpositional process [13]. More importantly, there seems to be little room left for an all-encompassing formal phonological structure as represented by the middle box in Fig. 6.

6. Concluding Remarks

With the advanced technological means we have today, there is no reason why tone research should not make a long stride in the next decade. To achieve that goal, however, scientific rigor needs to be widely applied to our research methodology and theoretical thinking. The principles outlined in this paper may serve only as an initial step.

References

[1] Banaji, M. R.; Crowder, R. G., 1989. The bankruptcy of everyday memory. *American Psychologist* 44: 1185-1193.

[2] Beckman, M. E., 1996. The parsing of prosody. *Language and Cognitive Processes* 11: 17-67.

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

[4] Chao, Y. R., 1922. Methods of experimental study of tone in Chinese speech. *Kexue [Science]* 7(9): 871-882.

[5] Chao, Y. R., 1968. *A Grammar of Spoken Chinese*. Berkeley, CA: University of California Press.

[6] Chen, Y. (2003). *The Phonetics and Phonology of Contrastive Focus in Standard Chinese.* Ph.D. dissertation. State University of New York at Stony Brook.

[7] Chen, Y.; Xu, Y., 2006. Production of weak elements in speech -- Evidence from f0 patterns of neutral tone in standard Chinese. *Phonetica* 63: 47-75.

[8] Cohen, A.; Collier, R.; 't Hart, J., 1982. Declination: Construct or intrinsic feature of speech pitch. *Phonetica* 39: 254-273.

[9] Cohen, A.; 't Hart, J., 1967. On the anatomy of intonation. *Lingua* 19: 177-192.

[10] Duanmu, S., 1994. Syllabic weight and syllable durations: A correlation between phonology and phonetics. *Phonology* 11: 1-24.

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

[12] Fant, G., 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.

[13] Fujisaki, H.; Wang, C.; Ohno, S.; Gu, W., 2005. Analysis and synthesis of fundamental frequency contours of Standard Chinese using the command–response model. *Speech communication* 47: 59-70.

[14] Gandour, J.; Potisuk, S.; Dechongkit, S., 1994. Tonal coarticulation in Thai. *Journal of Phonetics* 22: 477-492.

[15] Goldsmith, J. A., 1990. *Autosegmental and Metrical Phonology*. Oxford: Blackwell Publishers.

[16] Gordon, M., 2001. A typology of contour tone restrictions. *Studies in Language* 25: 405-444.

[17] Hombert, J.-M., 1974. Universals of downdrift: their phonetic basis and significance for a theory of tone. *Studies in African Linguistics* Supplement 5: 169-183.

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

[19] Honda, K., 1995. Laryngeal and extra-laryngeal mechanisms of F0 control. In *Producing Speech: Contemporary Issues — for Katherine Safford Harris*. F. Bell-Berti, & Raphael, L. (eds.) New York: AIP Press**:** 215-245.

[20] Honda, K.; Hirai, H.; S., M.; Shimada, Y., 1999. Role of vertical larynx movement and cervical lordosis in F0 Control. *Language and Speech* 42: 401-411.

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

[22] Howie, J. M., 1974. On the domain of tone in Mandarin. *Phonetica* 30: 129-148.

[23] Hyman, L.; Schuh, R., 1974. Universals of tone rules. *Linguistic Inquiry* 5: 81-115.

[24] Ishihara, S., 2002. Syntax-Phonology Interface of Wh-Constructions in Japanese. In *Proceedings of TCP 2002*, Tokyo. 165–189.

[25] Jiao, W., 2001. Zunhua fangyan kouyu zhong shangsheng de bianhua [Variations of the third tone in conversational speech of Zunhua dialect]. In *Proceedings of 5th National Conference on Modern Phonetics*, Beijing, China. 328-342.

[26] Jin, S. (1996). *An Acoustic Study of Sentence Stress in Mandarin Chinese.* Ph.D. dissertation. The Ohio State

University.
[27] Kelso, J. A. S., 1984. Phase transitions and critical behavior in human bimanual coordination. *American J. Physiology: Regulatory, Intergrative and Comparative* 246, R1000-R1004.
[28] Kelso, J. A. S.; Southard, D. L.; Goodman, D., 1979. On the nature of human interlimb coordination. *Science* 203: 1029-1031.
[29] Kuo, Y.-C.; Xu, Y.; Yip, M., in press. The phonetics and phonology of apparent cases of iterative tonal change in Standard Chinese. To appear in *Proceedings of TIE Conference*, Santorini.
[30] Ladd, D. R., 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
[31] Ladd, D. R.; Silverman, K. E. A., 1984. Vowel intrinsic pitch in connected speech. *Phonetica* 41: 31-40.
[32] Laniran, Y. O.; Clements, G. N., 2003. Downstep and high raising: interacting factors in Yoruba tone production. *Journal of Phonetics* 31: 203–250.
[33] Lee, T.; Kochanski, G. P.; Shih, C.; LI, Y. J., 2002. Modeling tones in continuous Cantonese speech. In *Proceedings of The 2002 International Conference on Spoken Language Processing*, Denver, CO. 2401-2404.
[34] Lehiste, I.; Peterson, G. E., 1961. Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America* 33, 419-425.
[35] Liberman, M.; Pierrehumbert, J., 1984. Intonational invariance under changes in pitch range and length. In *Language Sound Structure*. M. Aronoff and R. Oehrle. (eds.) Cambridge, Massachusetts: M.I.T. Press**:** 157-233.
[36] Liu, F.; Xu, Y., 2005. Parallel Encoding of Focus and Interrogative Meaning in Mandarin Intonation. *Phonetica* 62: 70-87.
[37] Mechsner, F.; Kerzel, D.; Knoblich, G.; Prinz, W., 2001. Perceptual basis of bimanual coordination. *Nature* 414: 69-73.
[38] Mixdorff, H., 2004. Quantitative tone and intonation modeling across languages. In Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing. pp. 137-142.
[39] Nakajima, S.; Allen, J. F., 1993. A study on prosody and discourse structure in cooperative dialogues. *Phonetica* 50: 197-210.
[40] Ohala, J. J., 1978. Production of tone. In *Tone: A linguistic survey*. V. A. Fromkin. (eds.) New York: Academic Press**:** 5-39.
[41] Ohala, J. J., 1990. There is no interface between phonology and phonetics: a personal view. *Journal of Phonetics* 18: 153-171.
[42] Peng, S.-h., 2000. Lexical versus 'phonological' representations of Mandarin Sandhi tones. In *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. M. B. Broe and J. B. Pierrehumbert. (eds.) Cambridge: Cambridge University Press**:** 152-167.
[43] Peterson, G. E.; Barney, H. L., 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24: 175-184.
[44] Pham, A. H., 2003. The Key Phonetic Properties of Vietnamese Tone: A Reassessment. In *Proceedings of The 15th International Congress of Phonetic Sciences*, Barcelona. 1703-1706.
[45] Pierrehumbert, J., 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation. MIT.
[46] Prince, A.; Smolensky, P., 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*: Rutgers University Center for Cognitive Science Technical Report 2.
[47] Prom-on, S.; Xu, Y.; Thipakorn1, B., 2006. Quantitative Target Approximation Model: Simulating Underlying Mechanisms of Tones and Intonations. In *Proceedings of The 31st International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France.
[48] Rump, H. H.; Collier, R., 1996. Focus conditions and the prominence of pitch-accented syllables. *Language and Speech* 39: 1-17.
[49] Shattuck-Hufnagel, S.; Turk, A. E., 1996. A Prosody Tutorial for Investigators of Auditory Sentence Processing. *Journal of Psycholinguistic Research* 25: 193-247.
[50] Shi, B., and Zhang, J., 1987. Vowel intrinsic pitch in Standard Chinese. In *Proceedings of The 11th International Congress of Phonetic Sciences*, Tallinn, Estonia. 142-145.
[51] Shih, C., 2000. A declination model of Mandarin Chinese. In *Intonation: Analysis, Modelling and Technology*. A. Botinis (ed.). Kluwer Academic Publishers**:** 243-268.
[52] Stevens, K. N., 1998. *Acoustic Phonetics*. Cambridge, MA: The MIT Press.
[53] Titze, I. R., 1994. *Principles of Voice Production*. New Jersey: Prentice Hall.
[54] van Santen, J. P. H.; Shih, C., 2000. Suprasegmental and segmental timing models in Mandarin Chinese and American English. *Journal of the Acoustical Society of America* 107: 1012-1026.
[55] Wang, B.; Xu, Y., 2006. Prosodic Encoding of Topic and Focus in Mandarin. In *Proceedings of Speech Prosody 2006*, Dresden, Germany.
[56] Wang, W. S.-Y.; Li, K.-P., 1967. Tone 3 in Pekinese. *journal of speech and Hearing research* 10: 629-636.
[57] Whalen, D. H.; Levitt, A. G., 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23, 349-366.
[58] Wong, Y. W., 2006. Contextual Tonal Variations and Pitch Targets in Cantonese. In *Proceedings of Speech Prosody 2006*, Dresden, Germany.
[59] Wong, Y. W., 2006. Realization of Cantonese Rising Tones under Different Speaking Rates. In *Proceedings of Speech Prosody 2006*, Dresden, Germany.
[60] www.phon.ucl.ac.uk/home/yi/downloads.html.
[61] www.phon.ucl.ac.uk/home/yi/f0_model.html.
[62] Xu, C. X.; Xu, Y., 2003. Effects of Consonant Aspiration on Mandarin Tones. *Journal of the International Phonetic Association* 33: 165-181.
[63] Xu, Y., 1994. Production and perception of coarticulated tones. *Journal of the Acoustical Society of America* 95, 2240-2253.
[64] Xu, Y., 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25, 61-83.
[65] Xu, Y., 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55, 179-203.
[66] Xu, Y., 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 27, 55-105.
[67] Xu, Y., 2001. Fundamental frequency peak delay in Mandarin. *Phonetica* 58, 26-52.
[68] Xu, Y., 2002. Articulatory constraints and tonal alignment. In *Proceedings of The 1st International Conference on Speech Prosody*, Aix-en-Provence, France. 91-100.
[69] Xu, Y., 2004. Understanding tone from the perspective of

production and perception. *Language and Linguistics* 5: 757-797.

[70] Xu, Y., 2005. Speech Melody as Articulatorily Implemented Communicative Functions. *Speech Communication* 46: 220-251.

[71] Xu, Y.; Liu, F., in press. Tonal alignment, syllable structure and coarticulation: Toward an integrated model. To appear in *Italian Journal of Linguistics*.

[72] Xu, Y.; Sun, X., 2002. Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America* 111: 1399-1413.

[73] Xu, Y.; Wallace, A., 2004. Multiple effects of consonant manner of articulation and intonation type on F0 in English. *Journal of the Acoustical Society of America* 115, Pt. 2: 2397.

[74] Xu, Y.; Wang, M., 2005. Tonal and durational variations as phonetic coding for syllable grouping. *Journal of the Acoustical Society of America* 117: 2573.

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

[76] Xu, Y.; Xu, C. X., 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33: 159-197.

[77] Xu, Y.; Xu, C. X.; Sun, X., 2004. On the Temporal Domain of Focus. In Proceedings of International Conference on Speech Prosody 2004, Nara, Japan. pp. 81-84.

[78] Yip, M., 2002. *Tone*. Cambridge: Cambridge University Press.

[79] Zemlin, W. R., 1988. *Speech and Hearing Science — Anatomy and Physiology*. Englewood Cliffs, New Jersey: Prentice Hall.

[80] Zhang, J., 2002. *The Effects of Duration and Sonority on Contour Tone Distribution -- A Typological Survey and Formal Analysis*. New York: Horn, Laurence.

[81] Zheng, X., 2006. Voice Quality Variation with Tone and Focus in Mandarin. This volume.