Maximum speed of pitch change and how it may relate to speech[a)]

Yi Xu[b)] and Xuejing Sun
Department of Communication Sciences and Disorders, 2299 North Campus Drive, Northwestern University, Evanston, Illinois 60208

(Received 23 May 2001; revised 13 November 2001; accepted 29 November 2001)

How fast speakers can change pitch voluntarily is potentially an important articulatory constraint for speech production. Previous attempts at assessing the maximum speed of pitch change have helped improve understanding of certain aspects of pitch production in speech. However, since only ''response time''—time needed to complete the middle 75% of a pitch shift—was measured in previous studies, direct comparisons with speech data have been difficult. In the present study, a new experimental paradigm was adopted in which subjects produced rapid successions of pitch shifts by imitating synthesized model pitch undulation patterns. This permitted the measurement of the duration of entire pitch shifts. Native speakers of English and Mandarin participated as subjects. The speed of pitch change was measured both in terms of response time and excursion time—time needed to complete the entire pitch shift. Results show that excursion time is nearly twice as long as response time. This suggests that physiological limitation on the speed of pitch movement is greater than has been recognized. Also, it is found that the maximum speed of pitch change varies quite linearly with excursion size, and that it is different for pitch rises and falls. Comparisons of present data with data on speed of pitch change from studies of real speech found them to be largely comparable. This suggests that the maximum speed of pitch change is often approached in speech, and that the role of physiological constraints in determining the shape and alignment of F_0 contours in speech is probably greater than has been appreciated.
[DOI: 10.1121/1.1445789]

PACS numbers: 43.70.Aj, 43.70.Bk, 43.70.Jt [AL]

I. INTRODUCTION

Speech is produced by a biomechanical system that has various inherent limitations. Many of these limitations may play a role in shaping the acoustic signal generated in speech production. One of them is the maximum speed at which speakers can change their pitch voluntarily. The importance of this limitation has not been widely recognized among those who are interested in the patterns of fundamental frequency variations in speech, however. Part of the reason for this lack of appreciation is probably a general feeling that in speech we approach our biomechanic limits only occasionally. In recent years, however, there has been accumulating evidence that ''time pressure'' may play a part in determining the shape and alignment of certain F_0 contours in speech (Caspers and Heuven, 1993; Ladd, Mennen, and Schepman, 2000; Xu, 1998, 2001). Unless some kind of limit is reached or approached, of course, time pressure should not make much difference. Nevertheless, the exact role the maximum speed of pitch change may play in speech is far from clear. This is partly because the data we have obtained about the speed of pitch change are far from complete, and partly because we have yet to pinpoint any direct link between maximum speed of pitch change and actual variations of F_0 contours in speech.

The maximum speed of pitch change was studied in the 70's by Ohala and Ewan (1973) and Sundberg (1979). Both studies used similar methods. Subjects were asked to shift from one pitch level to another as fast as possible upon command (Sundberg, 1979).[1] Then, the speed of pitch change was assessed by measuring the response time—time used to complete the fastest portion (the middle 75%) of a pitch shift, as illustrated in Fig. 1. The term ''response'' was first used by Ohala and Ewan (1973) in the phrase ''the response characteristics of the larynx in voluntary pitch change.'' The definition of the measurement, namely, time corresponding to the middle 75% of the pitch change, was also first used by Ohala and Ewan (1973). Sundberg (1979) used response time in his paper to refer to this measurement. Response time therefore provides a measurement of the time it takes the subject to respond to the command (hand waving or light flash as described in his paper) for making the pitch shift. Several findings of these early studies are potentially important for speech. First, the average response time was found to be around 79, 85, and 101 ms for pitch rises of 4, 7, and 12 semitones, respectively, and about 68, 70, and 73 ms for pitch drops of 4, 7, and 12 semitones. Thus, the speed of pitch change seems to increase as the interval becomes larger. Second, as also indicated by the measured response time, pitch lowering was faster than pitch elevation. Third, female subjects was found to have shorter response times than male subjects.

The estimates obtained by Ohala and Ewan (1973) and Sundberg (1979) constitute a very important step toward a

[a)]Part of the preliminary results of this study were presented at the 6th International Conference on Spoken Language Processing, Beijing, 2000 (Xu and Sun, 2000).
[b)]Electronic mail: xuyi@northwestern.edu

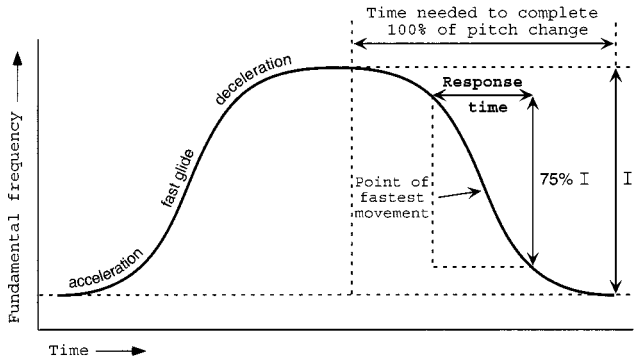

FIG. 1. First, illustration of *Response time*—time used to complete 75% of a pitch change, as defined by Ohala and Ewan (1973) and Sundberg (1979) (partially adapted from Sundberg, 1979). Second, illustration of the time needed to complete 100% of a pitch change. Third, conceptual division of a pitch movement into three phases: acceleration, fast glide, and deceleration.

clear understanding of the physiological limits on the speed of pitch change. However, those data are incomplete in two respects. First, as can be seen in Fig. 1, response time does not fully reflect the fastest instantaneous pitch movement possible, which should occur somewhere in the middle of the rising and falling ramps in the pitch change curve. Second, and more importantly, by definition, response time does not tell us how much time it takes for the speaker to complete 100% of a pitch shift, which is potentially much longer, as Fig. 1 clearly suggests.

The time needed to complete 100% of a pitch shift is potentially important for our understanding of pitch production in speech. In tone languages, for example, utterances often contain alternate high and low pitches. As observed in Xu (1997, 1999), the transitions between high and low pitches take a considerable amount of time, and the duration of the transition affects the shape of the F_0 contours. Furthermore, the minimum duration of the pitch transition may also limit how closely adjacent F_0 peaks or valleys can follow each other. Such limits may play a role in determining the alignment of F_0 peaks and valleys relative to segmental units such as syllables (Xu, 1999, 2001; Caspers and van Heuven, 1993).

It is also possible that articulatory limits on the speed of pitch change are never reached in speech, and that, instead, it is certain perceptual constraints that limit how fast pitch shifts are made in speech. This has been suggested by 't Hart, Collier, and Cohen (1990). In an effort to understand the rate of pitch change observed in speech, they considered both production and perception as possible contributing factors. Eventually, they dismissed production in favor of perception as the determining factor, based mainly on their interpretation of the data reported by Ohala and Ewan (1973) and Sundberg (1979). They first computed the maximum speed of pitch change using the response time for the 12-st condition as reported by Sundberg (1979), and came up with the value of 120 st/s. They then looked for the fastest pitch movement in Dutch and found it to be only 50 st/s. Based on this comparison, they concluded that articulatory limits simply could not have been responsible for the observed rate of pitch change in speech. Instead, they theorized, it must have been listeners' limited perceptual ability to distinguish between different rates of pitch change that has forced speakers to use slower pitch change rates in speech (pp. 71–75). A similar interpretation of Sundberg's (1979) data was also adopted by Caspers and van Heuven (1993).

Such interpretation of the data reported by Ohala and Ewan (1973) and Sundberg (1979) seems to have exaggerated the maximum speed of pitch change in two ways. First, the data from both studies indicate that the rate of pitch change is faster for a larger pitch interval than for a smaller one. So, unless a particular pitch movement actually spans 12 st, it is inappropriate to use the data for that interval as the direct indicator of speed of pitch change at other intervals. Second, response time, by definition (Sundberg, 1979), measures the time it takes to complete only 75% of the pitch change. Although it is not yet clear at this point how long it will take to complete 100% of a pitch change, it should certainly take longer than response time, as indicated by Fig. 1. Thus, it is inappropriate to treat response time as the ultimate indicator of the physiological limits on the speed of pitch change, because it corresponds to only part of the pitch shift.

A better way to estimate the actual physiological limits on the speed of pitch change would be to examine how long it takes to complete 100% of a pitch change at various pitch intervals. It probably would have been difficult, however, for Ohala and Ewan (1973) and Sundberg (1979) to measure the time interval of complete pitch changes with the experimental paradigm they employed, even if that was what they had in mind initially. In Fig. 1, at the beginning of a pitch shift, the F_0 movement seems to continually accelerate: changing very slowly at first, and gradually reaching full speed. Near the end of the pitch shift, the speed of movement decelerates, and the curve levels off gradually as the target pitch is being reached. To measure the time needed to complete 100% of the pitch shift as shown in Fig. 1, they would have to locate the exact points in time when the shift began and when it ended. However, the asymptotes near the onset and offset of the pitch shift make it hard to determine these points. To lessen the uncertainty in determining the end points of a pitch shift, therefore, it is necessary to minimize the duration of these asymptotes. One way to do that is to have the speaker produce a very quick succession of high and low pitches. This would generate a pitch undulation pattern that goes up and down rapidly, in which the lingering time on each pitch should be reduced to a minimum.

In the present study, we adopted a new paradigm to assess the maximum speed of pitch change by having subjects produce rapid pitch undulation patterns through imitation of rapidly alternating high–low pitch sequences. The maximum speed of pitch change is then estimated by measuring, among other things, the time used to complete 100% of each pitch shift. To see if these measurements can provide more direct indications as to whether and how physiological limits may play a critical role in shaping the F_0 contours of speech, we compared our data with data on speed of pitch change in real speech obtained in previous studies.

In addition, we also examined the effects of language, gender, pitch carrier (sustained schwa versus /malamalama/), and pitch shift direction on various measurements. To check

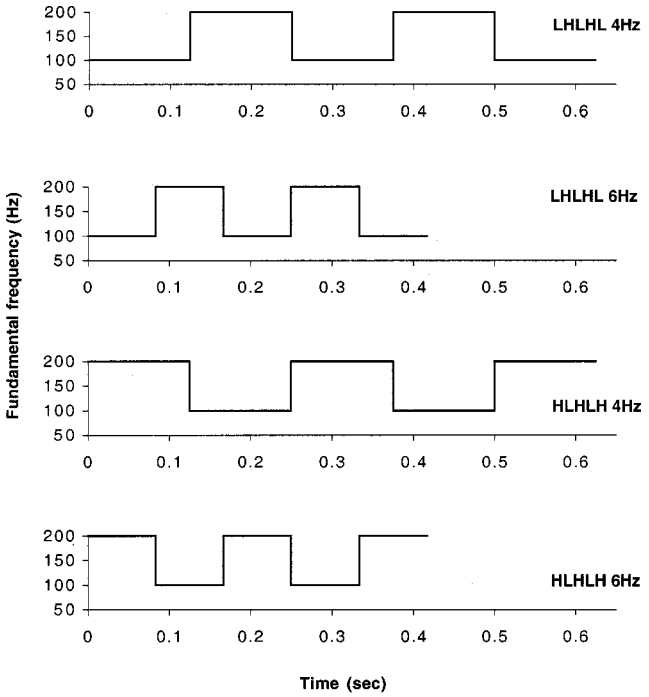

FIG. 2. Schematic representations of model pitch undulation patterns used as stimuli in the experiment, where 4 and 6 Hz refer to 4 and 6 undulation cycles per s, respectively. These examples have the base frequency of 100 Hz.

for possible differences due to language, we used native speakers of Mandarin and English as subjects. Mandarin has lexical tones (pitch patterns that can differentiate words) while English only has pitch accents related to intonation and word stress. It is possible that because successive high- and low-tone sequences frequently occur in Mandarin, native speakers of Mandarin can make pitch undulations faster than native speakers of English.[2] In addition, because pitch contours are usually carried by the vocalic portion of syllables in speech, it is possible that the syllable structure may either hinder the production of pitch undulation or facilitate it. To test which is the case, we used both CV syllables and simple vowels as the pitch carriers. To verify the gender difference as reported by Sundberg (1979), we also compared data from male and female speakers.

II. METHOD

A. Stimuli

The stimuli were model pitch undulation patterns to be imitated by subjects, as illustrated in Fig. 2. As can be seen in Fig. 2, the pitch undulation patterns are either HLHLH or LHLHL, where H and L represent relatively high and low pitches, respectively. The model pitch undulation patterns were built by modifying naturally produced voice samples. First, schwa-like vowels with relatively steady fundamental frequencies were recorded by a male and a female speaker. The duration of the vowel was approximately 1 s. Based on these original voice samples, a series of pitch manipulations was carried out to generate stimuli with desired pitch variation patterns.

Using the TD-PSOLA resynthesis function of the PRAAT program,[3] the fundamental frequencies of the original vowels were modified to generate a number of steady-state vowels. Based on the male voice, vowels with the base frequencies of 50, 60, 70, 80, 90, 100, and 115 Hz were generated. Based on the female voice, vowels with the frequencies of 115, 130, 145, 165, 185, 205, and 230 Hz were generated. These two ranges of frequencies correspond to the lower limits of male and female voices.

For each vowel at a particular fundamental frequency, 12 model pitch undulation patterns were generated. These pitch patterns differed in three ways. In terms of pitch variation pattern, they were either /HLHLH/ or /LHLHL/. In terms of pitch variation interval, the difference between H and L was, following Sundberg (1979), 4, 7, or 12 semitones. In terms of pitch variation rate, the duration of each HLH or LHL cycle in a pattern was either 1/4 or 1/6 s. In other words, the undulation frequency (i.e., the number of HLH or LHL cycles per second) of a model pitch pattern was either 4 or 6 Hz. The 4-Hz condition was to help warm up the subject during each trial, while the 6-Hz condition was to elicit the fastest pitch changes possible. It has been shown that 6 Hz is at the slower end of the involuntary vibrato rate in singing (Prame, 1994; Dejonckere, Hirano, and Sundberg, 1995). Presumably, *voluntary* pitch undulations are unlikely to exceed 6 Hz.

B. Subjects

Nineteen native speakers of American English (10 females and 9 males) and 22 native speakers of Mandarin Chinese (12 females and 10 males) between the ages of 18 and 45, recruited from Northwestern University campus, participated in the experiment. Subjects all reported having normal hearing, vision, and language ability. While some of the subjects had musical or voice training of some kind, none of them was a professional singer or involved in a professional singing group. Those with professional voice training were excluded so as to avoid the effect of extensive voice training on pitch undulation rate (Sundberg, 1979). The tasks of the experiment turned out to be too difficult for several subjects. Some of them were not able to produce the desired pitch shifts in the right order. Those subjects were all native speakers of English. The others produced pitch ranges smaller than two semitones in many trials. Those were both Mandarin and English subjects. As a result, only 36 subjects generated data suitable for analysis. Of the remaining subjects, 16 are English speakers (8 females and 8 males) and 20 are Chinese speakers (11 females and 9 males).

C. Procedure

The experiment was conducted in the Speech Acoustics Laboratory at Northwestern University. The subjects imitated the model pitch patterns using both a schwa and a syllable sequence (/malamalama/) as the pitch carriers. In total, each subject produced 240 trials (3 pitch intervals$\times$2 carriers$\times$2 patterns$\times$2 undulation rates$\times$2 sessions$\times$5 repetitions).

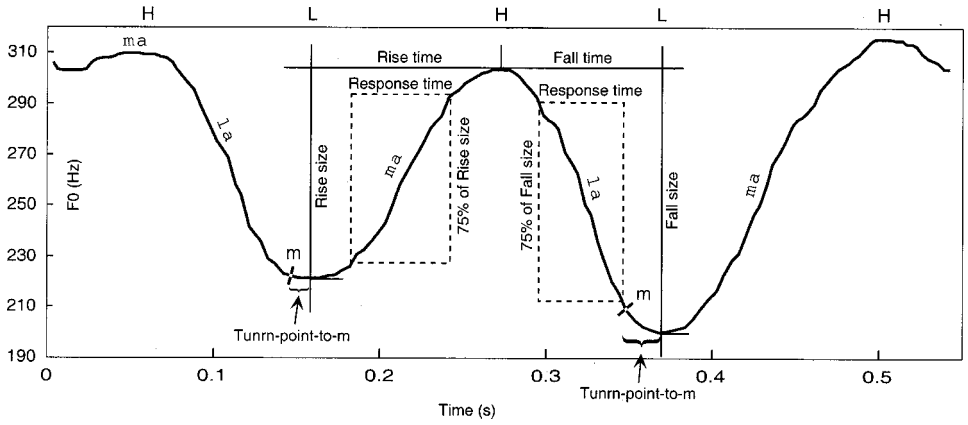

FIG. 3. Illustration of measurement of rise and fall excursion time, rise and fall response time, and turn-point-to-*m* in an HLHLH trial spoken with /malamalama/. The two cross points on the curve labeled "*m*" are the onsets of second and third /m/ in the trial. The fact that both turning points occur after "*m*" means that both turn-point-to-*m* values are negative.

During the experiment, the subject was seated in a sound-treated booth in front of a computer monitor. The experiment procedure was controlled by a set of HTML files, which were displayed by Netscape Navigator (Netscape Communications Corp.). A condenser microphone was used for the recording, and the vocalization was directly digitized onto a computer hard disk in a Macintosh G4 computer using SOUNDEDIT (Macromedia Inc.). For each subject, a comfortable pitch level was first determined before the start of the practice trials by choosing from a range of prerecorded voice samples played by the first HTML page. (Among the female subjects, seven selected the base frequency of 185 Hz, eight selected 205 Hz, and four selected 230 Hz. Among the male speakers, one selected the base frequency of 80 Hz, two selected 90 Hz, eight selected 100 Hz, three selected 115 Hz, and three selected 130 Hz.) The experimental stimuli were organized into three HTML pages, each containing model undulation patterns with the pitch intervals of 4, 7, or 12 semitones, respectively. On each page the undulation models are divided into two patterns—HLHLH and LHLHL, and two rates—4 and 6 Hz. The subject selected one of the stimuli each time by clicking on the corresponding button. The model pattern was then played through the loudspeaker. The subject was instructed to imitate the stimuli five times in each session, and as accurately as possible in terms of both pitch interval and undulation frequency.

The experimenter sat outside the recording booth, watching another computer screen showing the same display as seen by the subject, and listening to the subject's vocalization through a pair of headphones. The experimenter monitored the subject's performance and gave instructions when necessary. Since the task was somewhat difficult for some subjects, an intensive practice session was held for them before the real trials.

The whole recording process consisted of four sessions. In the first and third sessions, the subject imitated the pitch models with a schwa, while in the second and fourth sessions, the subject imitated the models with the syllable sequence /malamalama/. This particular syllable sequence was found in pilot tests to be the fastest vowel–sonorant sequence one could produce. Between sessions, the subject was given the chance to take a break. During each trial, i.e., for each model, the subject was allowed to replay the model pattern as many times as they wished. The experimenter would ask the subject to repeat a trial if it was felt to be necessary.

D. F_0 extraction and measurement

The F_0 extraction was done using a procedure similar to the ones used in Xu (1997, 1998, 1999, 2001). The procedure combines custom computer programming with ESPS/WAVES+ (Entropic Inc.). The digitized signals were transferred to a Dell workstation running on the LINUX platform. The *epochs* program was then used to mark every pitch period in each undulation sequence, 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 carefully for spurious vocal pulse markings such as double labeling and period skipping. Apparent errors were corrected manually.

While checking and correcting the vocal period labels, segmentation labels were also added at the onset and offset of all vocalizations and at the boundaries between /m/ and adjacent vowels for the /malamalama/ trials.

The vocal period and segment labels for each trial were saved in a text file. All the text files were then processed by a set of custom-written C programs. The programs converted the duration of pitch periods into F_0 values, and then smoothed the resulting F_0 curves using a *trimming algorithm* that eliminates sharp bumps and edges (Xu, 1999).[4] The trimmed F_0 curves were then subjected to further analysis using a set of custom-written MATLAB procedures. The following measurements were taken by the MATLAB procedures, most of which are illustrated in Fig. 3.

Excursion size (*rise* or *fall*)—pitch difference (in st) between adjacent F_0 minimum and maximum in the middle undulation cycle. Excursion size is expressed in semitone in order to make the data from individual speakers, especially across genders, more comparable.[5]

Excursion time (*rise* or *fall*)—time interval between adjacent F_0 maximum and minimum in the middle undulation cycle.

Excursion speed=excursion size/excursion time.

"*Response time*"—time interval corresponding to the middle 75% of *excursion size* (in Hz), as defined by Ohala and Ewan (1973) and Sundberg (1979).[6]

TABLE I. Mean values of various measurements under the effects of language, gender, direction (of pitch change), (pitch) carrier, and interval (of pitch change), together with probability values resulting from five-factor mixed-measure ANOVAs. Significant p values are printed in boldface.

	Language		Gender		Direction		Carrier		Interval		
	Chinese	English	Female	Male	Rise	Fall	Mala	Schwa	4	7	12
Excursion size (st)	4.4	5.8	5.0	5.0	4.8	5.3	5.1	5.0	3.8	4.7	6.6
	$p=$ **0.0037**		$p=0.8773$		$p<$ **0.0001**		$p=0.1396$			$p<$ **0.0001**	
Excursion time (ms)	125.3	141.2	128.7	136.4	132.5	132.2	133.6	131.2	125.7	128.2	143.2
	$p=$ **0.0219**		$p=0.2467$		$p=0.6987$		$p=0.2052$			$p<$ **0.0001**	
Response time (ms)	69.6	75.6	75.6	68.4	71.7	72.7	73.6	70.8	70.7	70.6	75.4
	$p=$ **0.0431**		$p=$ **0.0308**		$p=0.4608$		$p=$ **0.0208**			$p=$ **0.0005**	
Excursion speed (st/s)	35.9	42.1	40.1	37.0	36.5	40.8	38.6	38.7	30.8	37.5	47.6
	$p=0.1195$		$p=0.3674$		$p<$ **0.0001**		$p=0.8541$			$p<$ **0.0001**	
Maximum velocity (st/s)	60.8	72.4	65.0	67.1	61.3	70.6	66.4	65.6	50.3	62.4	85.1
	$p=0.0749$		$p=0.8323$		$p<$ **0.0001**		$p=0.3625$			$p<$ **0.0001**	

Maximum velocity—positive and negative extrema in the velocity curve corresponding to the rising and falling ramps in the middle undulation cycle. Velocity curves were computed by taking the first derivative of the F_0 curves after they were further smoothed by a five-point median filter and a seven-point (for male speakers) or 17-point (for female speakers) Hanning window.

For /malamalama/ files, the following measurements were also taken:

Peak-to-m—average time interval between the second and third F_0 maxima and onset of the second and third /m/ in LHLHL.

Valley-to-m—average time interval between the second and third F_0 minima and onset of the second and third /m/ in HLHLH.

Note that the value of *peak-to-m* or *valley-to-m* is negative if the peak or valley occurs after the onset of /m/, as illustrated in Fig. 3.

In the analyses, only data meeting the following criteria are included:

(a) Excursion size (rise or fall) >1 st;
(b) Excursion size (rise or fall) <2 standard deviations about the mean;
(c) Excursion time (rise or fall) <2 standard deviations about the mean.

Also, since the study is investigating the fastest speed of pitch change possible, only trials in the 6-Hz undulation frequency condition were processed for analysis. After applying these criteria to all trials in the 6-Hz condition, 3553 of the 4320 data points (82%) remained for further analysis. Of the excluded data points, 226 failed criterion (a), and 541 failed criteria (b) or (c).

III. ANALYSES

A. Effects of language, gender, direction of pitch change, pitch carrier, and interval of pitch change

Table I displays the mean *excursion size, excursion time, response time, excursion speed, and maximum velocity* broken down according to language (Chinese/English), gender (female/male), direction of pitch change (rise/fall), pitch carrier (malamalama/schwa), and interval of pitch change (4/7/12 st). Also displayed in the table are the probability values resulting from five-factor mixed-measure ANOVAs performed on the five measurements. Of the independent variables, language and gender are between-group factors, and the rest are within-group factors.

From Table I it can be seen that the effect of interval is significant for all measurements. Also, a set of Student–Newman–Keuls *post hoc* tests found the differences between all pairs of the three intervals to be significant at the 0.05 level, with the exception of excursion time and response time between 4 and 7 st. This indicates that (a) subjects managed to produce different excursion sizes for the three pitch-shift intervals, and (b) the speed of pitch change varied across the intervals. The mean interval sizes achieved by the subjects, however, are not quite what we had hoped for. In particular, for the 12-st condition, the mean interval achieved was only 6.5 st. Interestingly, the English subjects achieved greater excursion sizes than the Mandarin subjects.

In addition to excursion size, the effect of language is also significant for excursion time and response time. For both of them, the English subjects had greater means than the Mandarin subjects. This does not mean, however, that native English speakers are slower in making pitch changes. In fact, their excursion speed and maximum velocity are both somewhat faster than those of the Mandarin subjects, although neither difference reaches significance. It could be the case that the larger excursion size of the English subjects actually gave rise to the faster speed. This is partially verified by Table II, which shows that excursion speed and maximum velocity are highly correlated with excursion size, but not with excursion time and response time, despite the fact that time is actually in the equation for computing excursion speed. The fact that the English subjects produced larger pitch excursions and hence faster pitch changes than the Mandarin subjects is somewhat surprising to us, because presumably, speakers of a tone language should have better ability to make local pitch changes.[7]

As shown in Table I, the effect of direction is significant for excursion size, excursion speed, and maximum velocity. It is not significant, however, for excursion time and response time. While this is somewhat different from Sundberg (1979), where response time was found to be different for pitch lowering and pitch elevation, falls in the present data

TABLE II. Correlation (r) of various factors (computed from 432 mean values of each measurement: 2 directions×2 carriers×3 intervals×36 subjects=432 means).

	Excursion size	Excursion time	Excursion speed	Maximum velocity	Response time
Excursion size	1.000	0.384	0.859	0.920	0.212
Excursion time		1.000	−0.103	0.084	0.845
Excursion speed			1.000	0.956	−0.194
Maximum velocity				1.000	−0.106
Response time					1.000

are nevertheless consistently faster than rises.

There are also significant interactions between direction and interval for excursion size, excursion time, and response time. For excursion size, the interaction is largely due to greater differences between rise size and fall size at smaller intervals ($\Delta f = 0.6$ st and $\Delta f = 0.5$ st) than at the largest interval ($\Delta f = 0.2$ st). For excursion time and response time, on the other hand, the interaction is largely due to the lack of differences at the intervals of 4 and 7 st. In contrast, both excursion time and response time are longer when the interval is 12 st.

The main effect of carrier is significant only for response time, as shown in Table I. However, a number of interactions involving carrier reached or approached significance level. Interestingly, it is excursion size, excursion time, and response time that have significant or near-significant interactions, as can be seen in Fig. 4. English subjects show much larger differences between /malamalama/ and the schwa than Mandarin subjects for excursion size, excursion time, and response time. Furthermore, Mandarin speakers' excursion size is smaller when the carrier is /malamalama/ than when it is the schwa, whereas the difference with English speakers is reversed. It is possible that, for the Mandarin speakers, it is less natural to change pitch repeatedly within a sustained vowel than to associate each pitch value with a syllable, because the latter is similar to what they do in speaking their native language. If this is the case, in performing the task they may tend to use the usual pitch range for lexical tones which has been found to require only a small portion (up to 6 st, cf. Xu, 1999, in press) of a speaker's pitch range (up to 2 octaves, cf. Fairbanks, 1959). For the English subjects, in contrast, maybe associating a different pitch with each of the successive syllables is quite unnatural and consequently they had to use more effort in performing the task, resulting in a larger pitch range.[8]

Also, male subjects show larger differences between /malamalama/ and the schwa than female subjects for all three measurements (though only near-significance level for response time). The differences may seem to indicate that, with respect to these interactions, female speakers overall behave more like Mandarin speakers than like English speakers. However, the two probability levels do not seem high enough to warrant a clear conclusion about the gender effect at this point.

Somewhat surprisingly, the main effect of gender was not significant for any of the measurements except response time. But, response time is longer for female subjects than for male subjects, which is just the opposite of what is suggested by Sundberg's (1979) data. Nor was there any significant interaction between gender and direction as can be observed in Sundberg's data. As it turns out, however, there are other gender differences that are actually quite robust, as we will discuss next.

B. Excursion time versus response time

As suggested in Fig. 1, a complete pitch shift probably consists of three phases: acceleration, rapid glide, and deceleration.[9] Response time, as defined by Ohala and Ewan (1973) and Sundberg (1979), is the amount of time it takes the speaker to complete the middle 75% of a pitch change. Conceptually, therefore, pitch movement during response time corresponds largely to the fast glide phase of the pitch shift, and the rest of the excursion time to the initial acceleration and final deceleration. A comparison of excursion time with response time may thus let us see the distribution of time between the fast glide and acceleration–deceleration phases of a pitch shift. Table III displays excursion size, excursion time, response time, and the ratio of excursion time to response time broken down according to interval, direction, and gender.

As can be seen in the table, the ratio of excursion time to response time ranges from 1.62 to 2.07, and the mean ratio is 1.87. Thus, excursion time is always much longer than response time, and in many cases even more than twice as long. This indicates that a large portion of excursion time is missing when estimating the maximum speed of pitch change with response time alone.

Table III also reveals an interesting gender difference in terms of the ratio of excursion to response time. A five-factor (language, gender, direction, carrier, and interval) mixed ANOVA finds the main effect of gender highly significant for this ratio: $f(1,32) = 187.26$, $p < 0.0001$. Also highly significant are the main effect of interval [$f(1,32) = 24.24$, $p < 0.0001$] and the interaction between gender and interval [$f(1,32) = 26.63$, $p < 0.0001$]. As can be seen in the last row of Table III, for male speakers the excursion/response ratio remains fairly constant around 2. But, for female speakers the ratio increases steadily from 4 to 12 st, and is smaller than that of the males even for the largest intervals. Looking closer for the source of this difference, we notice in Table III that at each interval in both directions, female speakers have shorter excursion time but longer response time than male speakers. All this seems to indicate that female speakers use less time than male speakers in the acceleration and deceleration phases of the pitch shift. There are two possible ex-

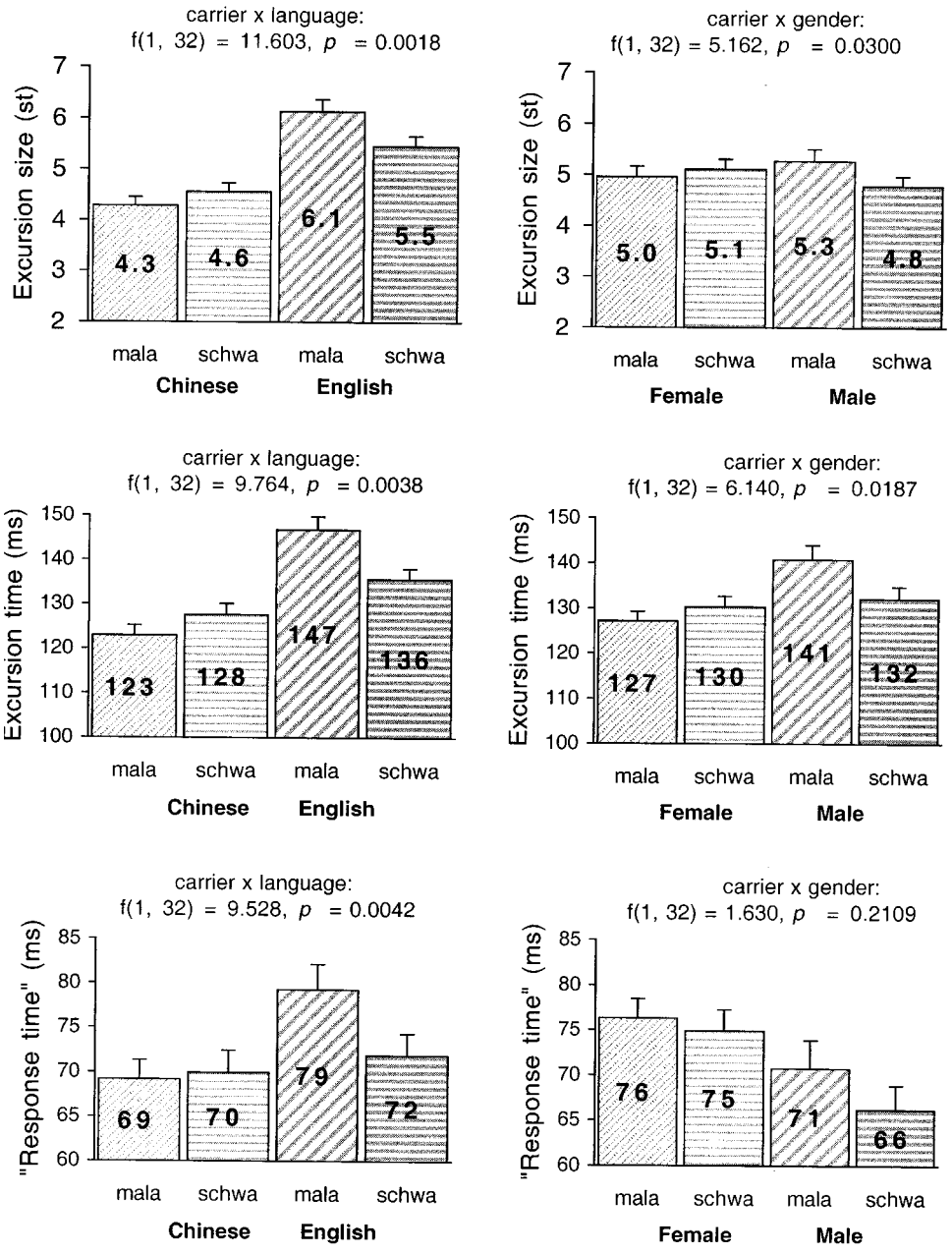

FIG. 4. Interactions of carrier with language and gender on excursion size, excursion time, and response time. Results of the five-factor mixed-measure ANOVAs for each interaction are displayed on top of the respective graph.

planations for this finding. The first is that female speakers have more powerful laryngeal muscles than male speakers so that they can start and stop a pitch shift faster than male speakers can. The second explanation, which we think is more plausible, is that female speakers have less laryngeal mass and hence less laryngeal inertia, thus needing less time than male speakers to initiate and end a pitch shift.

The main effects of direction and carrier also reach significance level [$f(1,32)=6.088, p=0.0191$; $f(1,32)=5.475, p=0.0257$]. But, the differences in the means are

TABLE III. Excursion size, excursion time, response time, and ratio of excursion time to response time. The ratios are means of individual ratios from all subjects.

Direction	Rise						Fall						
Interval	4		7		12		4		7		12		
Gender	f	m	f	m	f	m	f	m	f	m	f	m	Mean
Excursion size (st)	3.3	3.7	4.4	4.5	6.8	6.1	3.9	4.3	4.9	5.1	6.8	6.5	5.0
Excursion time (ms)	120	129	123	133	144	148	122	132	126	133	137	144	133
Response time (ms)	75	64	72	65	80	73	76	67	75	69	76	73	72
Excursion/response	1.62	2.07	1.70	2.07	1.82	2.04	1.62	2.00	1.69	1.96	1.83	2.02	1.87

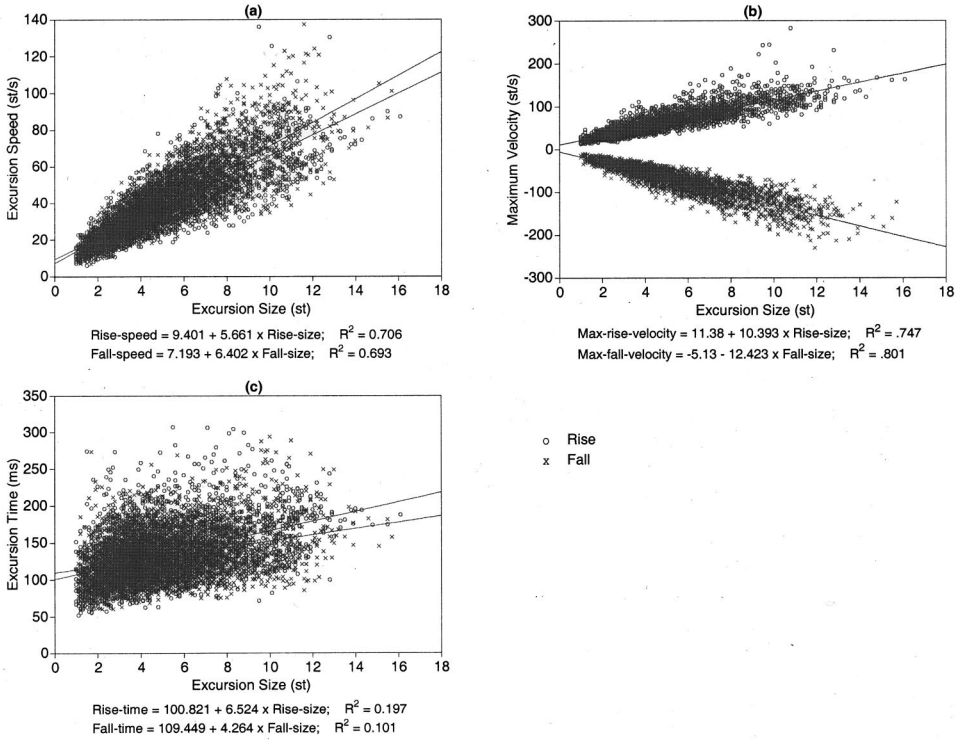

FIG. 5. Scatter plots of excursion speed, maximum velocity, and excursion time produced by all subjects as functions of excursion size. Only data points meeting the criteria listed in Sec. II D "F_0 extraction and measurement" are included. Each plot thus consists of 3553 data points.

so small (1.878 vs 1.844 for rise and fall, and 1.847 vs 1.876 for /malamalama/ and schwa) that we do not want to attribute them much importance.

C. Excursion time and excursion speed as functions of excursion size

We can see in Table I that excursion time, response time, excursion speed, and maximum velocity all vary significantly with pitch change interval. To observe their relation with pitch change interval in more detail, excursion time, excursion speed, and maximum velocity are plotted as functions of excursion size in all conditions for all the subjects in Fig. 5. As can be seen in the figure, in general both excursion speed and maximum velocity change fairly linearly with excursion size. Based on this observation, simple linear regressions were performed for each subject on both pitch rises and falls, with excursion size as independent variable and excursion time, excursion speed and maximum velocity as dependent variables.

Table IV shows the coefficients (intercept and slope, i.e., b and a in $z=b+ax$) of simple linear regressions of excursion time over excursion size, as well as excursion time computed with these regression parameters for the excursion sizes of 4, 7, and 12 st. Also shown in the table are the maximum, minimum, and standard deviation of the coefficients and computed excursion time across subjects. As can be seen, the deviation of the computed excursion time is quite large, especially when the excursion size is large. To reduce the deviation, three subjects (Nos. 11, 15, 36) whose computed excursion time at any of the six excursion sizes (3 sizes$\times$2 directions) deviates more than 2 standard deviations about the mean were taken out, and a new set of regression coefficients and computed excursion time values was obtained, as shown in Table V.

Regression coefficients for excursion speed and maximum velocity were also computed and are displayed in Tables VI and VII together with the predicted values at excursion sizes of 4, 7, and 12 st. Consistent with Fig. 5, for both excursion speed and maximum velocity, the mean R^2 values for pitch rise (0.723 and 0.734) and pitch fall (0.708 and 0.753) are quite high in Tables VI and VII. This further suggests that the relationship between excursion size and excursion speed is quite linear. The linear equations displayed in Tables VI and VII can therefore be used to predict the

TABLE IV. Coefficients (intercept and slope) of simple linear regressions of excursion time over excursion size, excursion time computed with these coefficients for the excursion sizes of 4, 7, and 12 st, and R^2 of the regression analyses. Data rows 1–4 display the mean, maximum, minimum, and standard deviation, respectively. In the Max and Min rows, all the values except R^2 are taken from the subject whose average rise time or fall time across the three intervals is the largest among all subjects. This way the maximum and minimum rise time can be computed with the corresponding intercept and slope values.

	Rise time (ms)						Fall time (ms)					
	Intercept	Slope	4 st	7 st	12 st	R^2	Intercept	Slope	4 st	7 st	12 st	R^2
Mean	90.4	9.7	129	158	206	0.355	100.9	6.3	126	145	177	0.201
Max	146.7	32.3	216	313	475	0.714	138.9	22.5	171	209	321	0.672
Min	48.7	3.1	91	105	128	0.053	51.3	−0.5	95	99	97	0.008
Std dev.	22.0	6.0	25.6	39.1	66.4	0.172	18.7	4.4	17.9	26.5	45.9	0.167

TABLE V. Coefficients (intercept and slope) of simple linear regressions of excursion time over excursion size, excursion time computed with these coefficients for the excursion sizes of 4, 7, and 12 st, and R^2 of the regression analyses (data from 33 subjects). Data rows 1–4 display the mean, maximum, minimum, and standard deviation, respectively. In the Max and Min rows, all the values except R^2 are taken from the subject whose average rise time or fall time across the three intervals is the largest among all subjects. This way the maximum and minimum rise time can be computed with the corresponding intercept and slope values in the table.

	Rise time (ms)						Fall time (ms)					
	Intercept	Slope	4 st	7 st	12 st	R^2	Intercept	Slope	4 st	7 st	12 st	R^2
Mean	89.6	8.7	124	150	194	0.356	100.4	5.8	124	141	170	0.193
Max	75.8	20.6	158	220	323	0.714	104.9	11.7	152	187	245	0.672
Min	73.1	4.6	91	105	128	0.053	102.1	−0.5	100	99	97	0.008
Std dev.	20.3	4.4	18.8	26.6	45.4	0.179	15.4	3.5	15.9	23.0	38.5	0.158

TABLE VI. Coefficients (intercept and slope) of simple linear regressions of excursion speed over excursion size, excursion speed computed with these coefficients for the excursion sizes of 4, 7, and 12 st, and R^2 of the regression analyses (data from 33 subjects). Data rows 1–4 display the mean, maximum, minimum, and standard deviation, respectively. In the Max and Min rows, all the values except R^2 are taken from the subject whose average rise time or fall time across the three intervals is the largest among all subjects. This way the maximum and minimum rise speed can be computed with the corresponding intercept and slope values in the table.

	Rise speed (st/s)						Fall speed (st/s)					
	Intercept	Slope	4 st	7 st	12 st	R^2	Intercept	Slope	4 st	7 st	12 st	R^2
Mean	10.8	5.6	33	50	78	0.723	8.9	6.2	34	52	83	0.708
Max	9.0	8.6	43	69	112	0.914	−0.6	10.3	41	72	123	0.906
Min	10.2	3.6	25	35	54	0.325	13.0	3.4	27	37	54	0.371
Std dev.	4.3	1.3	4.8	8.0	14.2	0.139	5.0	1.6	4.6	8.4	15.8	1.46

TABLE VII. Coefficients (intercept and slope) of simple linear regressions of maximum velocity over excursion size, maximum velocity computed with these coefficients for the excursion sizes of 4, 7, and 12 st, and R^2 of the regression analyses (data from 33 subjects). Data rows 1–4 display the mean, maximum, minimum, and standard deviation, respectively. In the Max and Min rows, all the values except R^2 are taken from the subject whose average rise time or fall time across the three intervals is the largest among all subjects. This way the maximum and minimum values can be computed with the corresponding intercept and slope values in the table.

	Max rise velocity (st/s)						Max fall velocity (st/s)					
	Intercept	Slope	4 st	7 st	12 st	R^2	Intercept	Slope	4 st	7 st	12 st	R^2
Mean	12.4	10.5	54	86	139	0.734	−6.8	−12.1	−55	−92	−152	0.753
Max	0.2	18.0	72	126	216	0.927	−11.8	−8.7	−46	−72	−116	0.941
Min	20.8	6.2	46	64	95	0.270	13.9	−18.2	−59	−113	−204	0.410
Std dev.	7.7	2.5	7.9	13.9	25.6	0.137	9.0	2.1	6.1	10.0	19.7	0.126

TABLE VIII. Mean values of turn-point-to-*m* (ms) under the effects of language, gender, turn type, and interval, together with probability values resulting from four-factor mixed-measure ANOVAs.

Language ($p = 0.0021$)	Chinese				English			
Gender ($p = 0.0241$)	Female		Male		Female		Male	
Turn type ($p = 0.0012$)	Valley	Peak	Valley	Peak	Valley	Peak	Valley	Peak
	−3.7	5.8	−0.2	21.9	12.9	18.1	24.6	29.8

TABLE IX. Comparison of response time measured by Sundberg (1979) and computed response time based on the present data.

		Rise response time			Fall response time		
	Gender	4 st	7 st	12 st	4 st	7 st	12 st
Sundberg (1979)	Female	75	81	98	62	68	70
	Male	83	89	104	73	72	75
Present study	Female	73	80	91	75	78	84
	Male	63	78	104	65	73	86

maximum speed of pitch change at different pitch change intervals. The data on maximum velocity as of yet do not have any real speech data with which to compare. However, in situations where the onset or offset of a pitch change is hard to determine, and maximum instantaneous velocity is the only measurable speed indicator, data in Table VII may be used as useful reference.

D. Alignment of peak and valley with syllable

For trials with /malamalama/ as the carrier, analysis was also done on the alignment of F_0 peaks and valleys relative to syllable boundaries. Table VIII displays the mean values of turn-point-to-*m* (including both peak-to-*m* and valley-to-*m*), broken down according to language, gender, and turn type. Also displayed in the table are the probability values resulting from four-factor mixed-measure ANOVAs. Of the independent variables, language and gender are between-group factors, and turn type and interval are within-group factors. The effect of interval is not significant and thus is not included in the table. As illustrated in Fig. 3, a positive value of turn-point-to-*m* means that the F_0 peak or valley corresponding to a syllable is realized *before* the end of the syllable; a negative value means that the peak or valley is realized *after* the syllable offset; and a small value, whether positive or negative, means that the peak or valley is realized close to the end of the syllable. The values in Table VIII indicate that Mandarin subjects produced F_0 peaks and valleys closer to the syllable offset than English subjects, and that there is a greater tendency for their peaks or valleys to occur *after* the end of the syllable. The same tendencies can also be seen in female versus male subjects, and in pitch falls versus pitch rises (i.e., valley versus peak).

IV. DISCUSSION

The main goal of the present study was to assess the maximum speed of pitch change in such a way that the assessment is more relevant to our understanding of F_0 contour variations in speech. By measuring not only response time as done in earlier studies (Ohala and Ewan, 1973; Sundberg, 1979) but also excursion time—time used to complete 100% of a pitch shift, the data so obtained can be more readily compared to data collected in studies of F_0 contours in real speech, as will be done next. As will be discussed subsequently, these comparisons make it possible for us to consider, in more realistic terms than before, implications of the maximum speed of pitch change on our understanding of F_0 contour production in speech in general. Finally, there are still inadequacies in the present study which can be improved in future research, as will be considered briefly.

A. Comparison with previous studies

There are a number of studies that have collected data to which the current data can be compared. In particular, Sundberg (1979), Caspers and van Heuven (1993), Ladd *et al.* (1999), Ladd *et al.* (2000), and Xu (1999) will be considered.

1. Sundberg (1979)

As mentioned earlier, Sundberg (1979), following Ohala and Ewan (1973), measured the response time of pitch shift, defined as the time to complete 75% of the pitch shift interval. To compare his data with ours, we estimated the mean response time from Fig. 3 of Sundberg (1979) (because no actual numbers were reported in the paper) and displayed them in the upper two rows in Table IX. The lower two rows of Table IX display the mean response time computed from our current data for males and females, respectively.

As can be seen in Table IX, while the computed response time values for pitch falls from the present study are slightly longer than those of Sundberg's, they are somewhat shorter than Sundberg's for pitch rises. Overall, the computed response time in the present study is comparable to that of Sundberg's. This suggests that the excursion time for Sundberg's subjects, had it been possible to measure it, could have been similar to that of the present study.

Also, the difference in speed between pitch rises and falls reported by Sundberg (1979) is largely confirmed. The mechanism behind this difference, however, is still unclear. Judging from the fact that the difference seems to increase with the size of pitch change, it is possible that it is due to the different muscles involved in pitch lowering and elevation. The former probably mainly involves the cricothyroid and thyroarytenoid, and the latter the infrahyoid strap muscles (sternohyoid, thyrohyoid, and sternothyroid) (Erickson, 1993; Hallé, 1994; Honda, 1995), which are more powerful, but slower at small ranges (Honda, 1995).

2. Caspers and van Heuven (1993)

Caspers and van Heuven (1993) examined the effect of time pressure on the realization of pitch rises and falls related to accents in Dutch. They measured the excursion size and F_0 slope of pitch rises and falls associated with prenuclear accents. Table X displays the fastest speed of pitch rises and falls for the female speaker and male speaker reported by Caspers and van Heuven. Also displayed in Table X are the excursion speeds computed from our current data for the average and fastest male and female speakers, respectively. The fastest speaker is the one whose computed excursion speed averaged across 4, 7, and 12 semitone is the greatest among all female or male speakers for pitch rises or falls. As can be seen in the table, in making pitch rises, the female speaker in their study is a bit faster than even the fastest female subject in the present study; the male speaker is faster than our average male speaker but slower than our fastest male speaker. In making pitch falls, their female speaker is faster than our average female speaker but slower than our fastest female speaker; their male speaker is a bit slower than our fastest male subject but about as fast as our average male subject. Overall, therefore, their speakers are faster than our average speakers but somewhat slower than our fastest speakers. This indicates that their two speakers were probably speaking near the speed limit when making the fastest pitch changes.

As mentioned in the Introduction, however, because they used response time for the 12-st interval reported by Sundberg (1979) as the actual limit of speed of pitch change, Caspers and Heuven concluded that the slope of pitch rises and falls was well within the articulatory limits. Comparisons of their data with ours as shown in Table X and the

TABLE X. Comparison of fastest excursion speed in Caspers and van Heuven (1993) and computed excursion speed based on the present data. The excursion sizes are as listed in Table 3 of their paper for the female and male speakers, respectively. The excursion speed from our data was computed using the intercepts and slopes of the fastest as well as the average male and female subjects in the present study at the same excursion sizes as those in Table 3 of Caspers and van Heuven.

		Rise		Fall	
	Subject	Size (st)	Slope/speed (st/s)	Size (st)	Slope/speed (st/s)
Caspers and van Heuven (1993)	Female	6.7	72	10.1	77
	Male	7.8	66	9.3	59
Present study	Average female	6.7	49	10.1	69
	Fastest female	6.7	67	10.1	96
	Average male	7.8	53	9.3	61
	Fastest male	7.8	72	9.3	84

newly-understood Sundberg data both suggest that the speakers examined by Caspers and van Heuven (1993) probably approached their maximum speed of pitch change quite frequently.

3. Xu (1999)

This study examined F_0 variations in Mandarin under the effects of tone and focus. The F_0 analysis in the study was done with a similar procedure as used in the present study. However, no data on the speed of pitch change were reported in the published paper. To extract data on the speed of pitch change from the raw data obtained in the study, we wrote a new C program to make the measurements. The program locates the F_0 peaks and valleys at the edges of pitch movements the same way as in the present study, and then measures the excursion time and excursion speed. The individual values of excursion speed were regressed over excursion size as in the present study for each of the eight Mandarin subjects. The regression equations were then used to compute excursion speed at the excursion intervals of 4, 7, and 12 st.

Table XI displays the mean excursion speed computed for the three intervals for different tones together with the excursion speed values as shown in Table VI and the mean excursion speed values from the Mandarin subjects alone. The values of speed of pitch change were divided into two groups: those associated with static tones and those with dynamic tones. Static tones refer to the H (High) and L (Low) tones. F_0 movements occur in this group when the tones of two adjacent syllables differ at the syllable boundary, such as in HL, LH, etc. Dynamic tones refer to the R (Rising) and F (Falling) tones, to which pitch movements are presumably intrinsic (Xu and Wang, 2001).

As can be seen in Table XI, although excursion speed associated with the static tones in Xu (1999) is much slower than the maximum excursion speed obtained in the present study, the excursion speed associated with the dynamic tones is fairly comparable with the present data both from all the subjects and from the Mandarin subjects alone. This comparison is quite interesting, because it shows that in speech, the maximum speed of pitch change is approached only when there is a strong demand for a fast pitch change.

4. Ladd et al. (1999) and Ladd, Mennen, and Schepman (2000)

These two studies investigated the alignment of F_0 peaks and valleys in the prenuclear rising accent in English (Ladd *et al.*, 1999) and Dutch (Ladd *et al.*, 2000). We computed the speed of pitch change (st/s) from the data reported in these studies and listed them in Table XII. Also listed in the table is the speed of pitch change in the present study estimated using the coefficients in Table VI for the same excursion sizes.

From Table XII it can be seen that in both cases, the speed of pitch change in those two studies is somewhat slower than the estimated speed obtained in the present study. This is despite the fact that both studies include conditions where time pressure is potentially applied to the realization of pitch movements. We do notice one thing that is different about these two studies when compared to the studies discussed earlier, however. That is, the prenuclear accents

TABLE XI. Comparison of excursion speed estimated from Xu (1999) and from the present data (same as in the first row of Table VI).

		Rise speed			Fall speed		
		4 st	7 st	12 st	4 st	7 st	12 st
Xu (1999)	Static tone	24	37	58	21	35	57
	Dynamic tone	31	51	83	29	49	81
All speakers in present study		33	50	78	34	52	83
Mandarin speakers in present study		34	52	82	34	53	84

TABLE XII. Comparison of excursion speed in Ladd *et al.* (1999) and Ladd *et al.* (2000) with computed excursion speed based on our data (Table VI). For Ladd *et al.* (1999), the excursion sizes are computed from data in Appendix B of their paper, the excursion time is obtained from Fig. 3 and Table 2 of their paper. For Ladd *et al.* (2000), for their experiment 1 the rise size was estimated based on their Table I, and rise time was computed from data in Table I as well as in the text. For their experiment 2, the rise size is from endnote 2 and rise time from both their Table II and endnote 2. The rise speeds in both studies were calculated by dividing rise size with rise time.

	Rise size (st)	Rise speed (st/s)
Ladd *et al.*, 1999: Experiment 1 fast	3.7	21
Present study	3.7	31
Ladd *et al.*, 1999: Experiment 2 fast	3.4	21
Present study	3.4	29
Ladd *et al.*, 2000: Experiment 1 fast	5.4	23
Present study	5.4	40
Ladd *et al.*, 2000: Experiment 2	6.5	31
Present study	6.5	46

in these studies always occur on a syllable that is followed by an unstressed syllable. If we assume that an unstressed syllable either does not have a pitch target of its own, or carries only a rather weak pitch target, then it is possible that the rise in a prenuclear accent that precedes an unstressed syllable is not implemented under the greatest time pressure. In other words, they are somewhat similar to the situation of the static tones in Mandarin, whose implementation also does not seem to require maximum speed of pitch change, as has been shown in Table XI. Naturally, the validity of this interpretation awaits closer examination in future studies.

B. Implications

As suggested earlier, a rapid pitch shift should consist of three phases: acceleration, rapid glide, and deceleration. Previous studies of the maximum speed of pitch change seem to have focused mainly on the second phase, i.e., the rapid glide (Ohala and Ewan, 1973; Sundberg, 1979). The present study takes all three phases into consideration when estimating the speed of pitch change. As it turns out, excursion time is nearly twice as long as response time. Furthermore, it is found that the speed of pitch change varies quite linearly with the size of pitch change, and that it varies also with the direction of pitch change. These findings have many implications for our understanding of pitch contours in speech as well as other aspects of speech production. In the following, we will discuss just a few of these implications.

1. How often is the maximum speed of pitch change reached in speech?

The role of articulatory constraints has been widely recognized in the phonetics and phonology literature. However, rarely do we see serious discussion on whether limits on the speed of articulatory movements are actually reached. Perhaps this is because of the general belief that human beings as biological systems would not allow their physiological limits to be approached very often when performing a task as routine as speech. Instead, more consideration is given to the economy of effort, as defined by Lindblom (1982), as the ultimate constraint in speech production. Economy of effort implies that the speaker is capable of making a more extreme articulation but chooses not to. A physiological limit, on the other hand, is a threshold that the speaker simply cannot cross. One study that does seriously consider the possible role of articulatory limit on the speed of pitch change in determining various aspects of F_0 contours in speech is 't Hart *et al.* (1990). As discussed in the Introduction, however, their interpretation of the data reported by Sundberg (1979) underestimated the actual articulatory limits on the speed of pitch change. The comparison of present data with those of Xu (1999) discussed earlier suggests that pitch change speed comparable to that obtained with a paradigm as demanding as that employed in the present study can be easily observed in real speech in Mandarin. For Dutch, the full excursion size found by 't Hart *et al.* (1990) is around 6 st (p. 53). At this interval, the speed of 50 st/s they reported is also comparable to the excursion speed in Table VI at the same interval. Also, as shown in Table X, the fastest pitch change speed reported by Caspers and van Heuven (1993) is comparable to the maximum speed of pitch change at similar pitch shift intervals. Furthermore, as mentioned by 't Hart *et al.*, in English, full-size rises and falls can span an octave and the rate of change can reach 75 st/s (p. 49). This again is comparable to the computed mean excursion speed for the 12-st interval shown in Table VI. These comparisons seem to suggest that the maximum speed of pitch change is probably approached or even reached more often than we have realized.

Note that this does not mean that the maximum speed of pitch change is reached all the time. Rather, there are many situations in which the thresholds are not likely approached. For example, the production of the static tones (H, L) in Mandarin probably does not often require maximum speed of pitch change, as Table XI seems to suggest. Also, the production of the prenuclear accent in English and Dutch probably does not call for maximum speed of pitch change if the stressed syllable is followed by an unstressed syllable, as Table XII appears to suggest.[10] What seems critical is that in each specific situation we need to try to recognize if a particular biomechanical limit may be approached and whether a condition exists that necessitates the approximation of that limit. In the following we will discuss a number of such situations and examine how maximum speed of pitch change may play a role in shaping certain F_0 contours in speech.

2. How may contextual tonal variations relate to maximum speed of pitch change?

In a series of studies on contextual tonal variations in Mandarin (Xu, 1994, 1997, 1999), it was found that the F_0 contour of a tone varies extensively with the offset F_0 of the preceding tone, especially when there is no voiceless consonant separating the vowels. The H tone in Mandarin, for example, is produced with an apparent rising contour when it follows the L tone. Likewise, the L tone is produced with an apparent falling contour when preceded by the H tone. Xu and Wang (2001) suggest that these seemingly long transitions are due to the fact that it takes time to make the required pitch change when shifting from one tone to another. What was not known, however, was how much time it would take for such transitions to complete. As shown in Table V, it

would take on average 142 ms (computed with the mean intercept and slope for rise time in the table) for a Mandarin speaker to complete a 6-st pitch rise. This means that in a syllable with an average duration of 180 ms (Xu, 1999), the greater half of the F_0 contour in the syllable would have to be used for completing the pitch rise from the L tone to the H tone even if the maximum speed of pitch has been achieved. The long transitions found in Xu (1997, 1999) now seem to have a clearly plausible articulatory explanation: speakers probably have no way of avoiding them, given the limit of their laryngeal physiology.

In many African tone languages, e.g., Yoruba, the H tone is said to change into a rising tone when preceded by the L tone, and the L tone is said to change into a falling tone when preceded by the H tone (Hyman and Schuh, 1974). While there are various phonological accounts of this kind of tonal variation (e.g., Hyman and Schuh, 1974; Goldsmith, 1990; Manfredi, 1993), it is not yet clear if such changes are due to speaker's intentional change of the articulatory target associated with the tone. From the limited duration data that can be obtained from Laniran (1992) and Akinlabi and Liberman (1995), it seems that the average syllable duration in Yoruba is no longer than that in Mandarin. This suggests that these dynamic F_0 patterns in Yoruba probably have much to do with speakers' articulatory limitations. If the maximum speed of pitch change found in the present study is universal, the rise in the H tone and fall in the L tone are probably inevitable whenever they are preceded by a tone with a very different offset pitch.

3. What are the linguistically meaningful pitch targets and how are they realized in speech?

As has been observed in a number of recent studies, certain F_0 events such as F_0 peaks and valleys maintain a relatively stable alignment with the onset or offset of the syllable (Arvaniti, Ladd, and Mennen, 1998; Caspers and van Heuven, 1993; Kim, 1999; Ladd *et al.*, 2000; Prieto, v. Santen, and Hirshberg, 1995; Xu, 1998, 1999, 2001). There are disagreements, however, over the interpretation of these alignment patterns. In particular, Ladd and his colleagues argue that these patterns indicate that F_0 turning points are linguistically meaningful targets and are ''anchored'' by speakers at the onset or offset of the syllable, and that observed F_0 shapes are merely interpolations between these targets (Arvaniti *et al.*, 1998; Ladd *et al.*, 1999; Ladd *et al.*, 2000). An alternative view recently offered by Xu and Wang (2001) and Xu (in press) contends that observed F_0 events such as peaks and valleys are not necessarily the underlying functional units *per se*. Rather, they are mostly products of the interaction between underlying pitch targets and various articulatory constraints. For example, the H, L, R, and F tones in Mandarin probably have the underlying pitch targets [high], [low], [rise], and [fall]. In speech production, these targets are synchronously implemented with the syllables that carry them due to, presumably, the constraints of coordinated movements (Kelso, 1984; Schmidt, Carello, and Turvey, 1990). Due to the constraint of the maximum speed of pitch change, however, the realization of these targets in contexts often deviates much from their realization in isolation, resulting in contextual tonal variations as discussed in the previous section. Furthermore, according to this view, the occurrence of F_0 peaks and valleys requires the right sequence of pitch targets, and the alignment of the peaks and valleys depends on the properties of the pitch targets involved. For example, in an LHL sequence, F_0 has to rise from the first L tone to realize the [high] of the H tone, and then fall to realize the [low] of the second L tone. This will result in a rising F_0 contour during the H-carrying syllable, a falling contour during the second L-carrying syllable, and a peak near the boundary between the second and third syllables.

What the findings of the present study tell us is that, no matter what form the linguistically meaningful targets take, implementing them takes time. If, for example, the speaker's task is to anchor an F_0 minimum at the onset of a syllable-initial consonant, as suggested by Ladd and colleagues, an average speaker would have to start the F_0 movement toward this low point at least 124 ms earlier, even if the range of the movement is just 4 st (cf. Table V). Furthermore, the speaker would have to adjust the onset of a pitch movement according to the size of the F_0 excursion toward that low point. This has yet to be confirmed by empirical data.

If, instead, the speaker's task is to implement a pitch target such as [high] in synchrony with a syllable, as suggested by Xu and Wang (2001), there would be no need to anticipate the size of the F_0 movement toward this target. Rather, the speaker just needs to start the implementation of the pitch target at the onset of the syllable and end the implementation at the offset of the syllable. Because it takes at least 124 ms to raise or lower pitch by 4 st according to the present data (Table V), much of the earlier F_0 contour during a syllable would form a transition from the preceding pitch target toward the current target. Furthermore, the shape of the transition and the height and slope of the F_0 contour near the end of the syllable would all vary depending on the magnitude and direction of the difference between the two adjacent targets and the duration of the syllable. Additionally, depending on the duration of a syllable, an F_0 peak associated with it would occur either before or after its offset. All of these have indeed been found in Mandarin (Xu, 1997, 1998, 1999, 2001).

The present data therefore seem to provide some support for the view that, at least in Mandarin, underlying pitch targets such as [high], [low], [rise], and [fall] constitute part of the meaningful linguistic units and they are produced synchronously with their associated syllables. It is possible; however, languages like English and Dutch are very different from Mandarin in terms of underlying pitch targets and their alignment, and this difference may explain the contrast between the aforementioned views. Further studies are needed to resolve this issue.

C. Limitations, caveats, and future directions

Despite the significance of the data obtained in the present study, we are aware of their limitations, and we want to also point out a few caveats and identify possible future directions. First, from Tables IV–VII it is apparent that there are large variations across subjects. Part of the variability

may be due to subjects' different abilities to perform the arbitrary task of the experiment. For example, compared to Sundberg's (1979) nonsinger subjects who were all taking a musical class at the time of the experiment, ours had rather diverse musical backgrounds. Although we did not find any contribution of musical training, the lack of musical training of some subjects may have contributed to the difficulty they experienced while trying to perform the task.

Second, the 6-Hz undulation rate that we used for the stimuli is probably a bit too fast, since even our fastest subjects did not achieve that rate. We used 6 Hz to ensure that we get the fastest speed possible. But, it may also have contributed to the difficulty some subjects experienced during the experiment. The other source of the variability may be each individual subject's true idiosyncratic speed of pitch change. It would be interesting in future studies to examine whether individual speakers' F_0 contour patterns are directly linked to their own maximum speed of pitch change.

Third, the significant effects of language, gender, and turn type on turn-point-to-*m*, as shown in Table VIII, are somewhat puzzling to us. In the /malamalama/ condition, what the subjects were asked to do is to produce the pitch undulation patterns together with the syllable sequences. Although there were no explicit instructions as to how precisely they should align the two, the implied requirement is that they produce the two synchronously. The patterns shown in Table VIII, however, do not seem to fit what one might predict from previous data. We know that female speakers might have faster speed of pitch change according to Sundberg's (1979) data, or they may have faster F_0 movement acceleration and deceleration as suggested by our new data discussed above. We also know that Mandarin probably requires higher precision of pitch target alignment than English because virtually every syllable is specified with a lexical tone. Finally, we know from data reported by Ohala and Ewan (1973), Sundberg (1979), and the current study that lowering pitch is faster than raising it. As speculated by Ohala (1978), ''since they can be accomplished quicker, they [falling tones] might be less likely than rising tones to 'spill over' onto the next syllable'' (p. 31). All these seem to suggest that the turning points should more likely occur after rather than before the end of the pitch-associated syllable for male than for female speakers, for Mandarin than for English speakers, and for rises than for falls. The fact that the opposite of all of these was found in the present data might suggest that there is some subtle but critical difference between subjects' tasks in the present study and the task of producing linguistically meaningful words and phrases.

Finally, this study did not look directly into the physiology of pitch production. And, the few physiological studies we reviewed could not provide us with direct explanations about the speed of pitch change observed in this study. So, it is not yet clear to us what, at the muscular level, makes changing pitch take as much time as found in the present study. This, again, can be resolved only by future studies.

V. CONCLUSIONS

The goal of the present study was to assess the maximum speed of pitch change in such a way that the results can be directly compared to data from real speech. This was motivated by our realization that previous attempts at measuring the speed of pitch change have provided only a partial picture. They obtained data only on response time—time needed to complete 75% of a pitch shift. As shown in Fig. 1, that corresponds only to the fast glide phase of a pitch shift. It became apparent to us that the initial acceleration and final deceleration should also be taken into consideration before the data on the speed of pitch change can be fully useful for speech research. In this study, we adopted a new experimental paradigm in which subjects produced rapid pitch shifts by imitating a series of model pitch undulation patterns. This enabled us to measure the complete duration of each pitch shift as well as that of the response time as defined in previous studies.

As it turns out, it takes nearly twice as long to complete an entire pitch shift as it takes to execute the middle 75% of the shift. This finding indicates that physiological limitation on the speed of pitch movement is probably much greater than has been recognized. More interestingly, we find that the maximum speed of pitch change obtained in this study is comparable to the maximum speed of pitch change observed in a number of existing studies on real speech. This suggests that the role of physiological constraints in determining the shape and alignment of F_0 contours in speech is probably more important than has been appreciated. While it is likely that very often articulatory movements do not reach maximum speed due to speakers' choice of not using excessive effort, in many other occasions, as the new data show, the absolute limit may indeed have been approached or even reached. This means that, for pitch contour production at least, absolute articulatory limits, just as the economy of efforts as suggested by Lindblom (1982), probably constitute a major articulatory constraint that helps to shape the acoustic signal in speech.

Our data also demonstrate more clearly than before the linear relations between the *size* and *speed* and between *size* and *peak velocity* of pitch change. Similar linear relations have been found between peak velocity and movement amplitude in both speech and nonspeech movements (Hertrich and Ackermann, 1997; Löfqvist and Gracco, 1997). This suggests that F_0 movements are probably quite similar to other speech and nonspeech movements. The linear relation between the size and speed of pitch movement also suggests that it is imperative to know the pitch change interval when determining whether the maximum speed of pitch change is approached in a given case. Linear regression coefficients displayed in Tables V–VII can be used in future studies as references for determining for each pitch variation pattern in a language, how much of it is explainable in terms of articulatory constraints, and how much of it has to be explained by other mechanisms such as language specific phonological rules. The method of inducing rapid pitch undulation patterns from human subjects as used in the present study can also be employed in future studies to test various speaker groups as well as individual speakers to see whether and how their speech patterns are related to their idiosyncratic limits on the speed of pitch change.

Finally, although our findings are about the speed of

changing pitch, they also raise the question of whether the maximum speed of other articulatory movements is also approached in speech more frequently than has been recognized. Future efforts in finding answers to this question may help further our understanding of speech production and speech perception in general.

ACKNOWLEDGMENTS

This work is supported in part by NIH Grant No. DC03902. We would like to thank Bob Ladd, two anonymous reviewers, and Yiya Chen for their valuable comments and suggestions on earlier versions of this paper.

[1]No information was given as to what vowel(s) the subjects used to sing the pitch notes in the study.

[2]One reviewer cautioned us that English may actually produce faster pitch change than Chinese, because prominence is realized in English as F_0 contrast between adjacent syllables. As found in Xu (1999), however, this kind of contrast is used just as readily in Mandarin. In any case, as it turned out, we found no significant difference in speed of pitch change between speakers of these two languages.

[3]We thank Paul Boersma and the Praat project at Institute of Phonetic Sciences, University of Amsterdam, for making their program freely available to speech researchers.

[4]As described in Xu (1999), the trimming algorithm removes only local bumps in an F_0 curve involving adjacent vocal periods. As shown in Fig. 2 in Xu (1999), the ''trimming'' makes sure that sharp local bumps are not mistaken as true extreme pitch points by subsequent algorithms that locate peaks and valleys in an F_0 curve.

[5]We also considered the ERB scale (Hermes and van Gestel, 1991; Hermes and Rump, 1994), but did not find it appropriate for our purpose, because it reduces the speaker differences too heavily.

[6]Although it is conceptually more appropriate to calculate response time using excursion size expressed in semitone, numerically the difference between the two ways of measurement is quite small. Besides, it is critical that our data are in a form that is directly comparable to that of the previous studies.

[7]We also examined the possible contribution of musical training, but did not find any.

[8]It is also possible that it is the phonetic difference in [1] in the two languages that caused the difference in excursion time and response time. The American English [1] is known to be ''dark,'' i.e., with the back of the tongue as well as the tongue tip raised, whereas the Mandarin [1] is quite ''light.'' The back of the tongue is bulkier than the tip of the tongue. Raising it may require more time and it may have lengthened the entire syllable.

[9]We are not suggesting, however, that there are clear boundaries between the phases. Rather, the F_0 contour throughout a pitch shift is continuous, and the division among the three phases is conceptual and arbitrary.

[10]The same can also be seen in the rise data in Table 3 in Caspers and van Heuven (1993).

Akinlabi, A., and Liberman, M. (**1995**). ''On the phonetic interpretation of the Yoruba tonal system,'' Proceedings of The 13th International Congress of Phonetic Sciences, Stockholm, pp. 42–45.

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

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

Dejonckere, P. H., Hirano, M., and Sundberg, J. (**1995**). *Vibrato* (Singular, San Diego).

Erickson, D. (**1993**). ''Laryngeal muscle activity in connection with Thai tones,'' Ann. Bull. Res. Inst. Logoped. Phoniatr. **27**, 135–149.

Fairbanks, G. (**1959**). *Voice and Articulation Drillbook* (Harper & Row, New York).

Goldsmith, J. A. (**1990**). *Autosegmental and Metrical Phonology* (Blackwell, Oxford).

Hallé, P. A. (**1994**). ''Evidence for tone-specific activity of the sternohyoid muscle in Modern Standard Chinese,'' Lang. Speech **37**, 103–123.

Hermes, D., and van Gestel, J. (**1991**). ''The frequency scale of speech intonation,'' J. Acoust. Soc. Am. **90**, 97–102.

Hermes, D. J., and Rump, H. H. (**1994**). ''Perception of prominence in speech intonation induced by rising and falling pitch movements,'' J. Acoust. Soc. Am. **96**, 83–92.

Hertrich, I., and Ackermann, H. (**1997**). ''Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures,'' J. Acoust. Soc. Am. **102**, 523–536.

Honda, K. (**1995**). ''Laryngeal and extra-laryngeal mechanisms of *F*0 control,'' in *Producing Speech: Contemporary Issues—for Katherine Safford Harris*, edited by F. Bell-Berti and L. Raphael (AIP, New York), pp. 215–245.

Hyman, L., and Schuh, R. (**1974**). ''Universals of tone rules,'' Ling. Inq. **5**, 81–115.

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

Kim, S.-A. (**1999**). ''Positional effect on tonal alternation in Chichewa: Phonological rule vs phonetic timing,'' Proceedings of Annual Meeting of Chicago Linguistic Society, Chicago, pp. 245–257.

Ladd, D. R., Faulkner, D., Faulkner, H., and Schepman, A. (**1999**). ''Constant 'segmental anchoring' of *F*0 movements under changes in speech rate,'' J. Acoust. Soc. Am. **106**, 1543–1554.

Ladd, D. R., Mennen, I., and Schepman, A. (**2000**). ''Phonological conditioning of peak alignment in rising pitch accents in Dutch,'' J. Acoust. Soc. Am. **107**, 2685–2696.

Laniran, Y. (**1992**). ''Intonation in Tone Languages: The Phonetic Implementation of Tones in Yorùbá,'' Ph.D. dissertation, Cornell University.

Lindblom, B. (**1982**). ''Economy of speech gestures,'' in *The Production of Speech*, edited by P. F. MacNeilage (Springer, New York), pp. 217–245.

Löfqvist, A., and Gracco, V. (**1997**). ''Lip and jaw kinematics in bilabial stop consonant production,'' J. Speech Hear. Res. **40**, 877–893.

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

Ohala, J. J. (**1978**). ''Production of tone,'' in *Tone: A Linguistic Survey*, edited by V. A. Fromkin (Academic, New York), pp. 5–39.

Ohala, J. J., and Ewan, W. G. (**1973**). ''Speed of pitch change,'' J. Acoust. Soc. Am. **53**, 345(A).

Prame, E. (**1994**). ''Measurements of the vibrato rate of ten singers,'' J. Acoust. Soc. Am. **96**, 1979–1984.

Prieto, P., v. Santen, J., and Hirshberg, J. (**1995**). ''Tonal alignment patterns in Spanish,'' J. Phonetics **23**, 429–451.

Schmidt, R. C., Carello, C., and Turvey, M. T. (**1990**). ''Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people,'' J. Exp. Psychol. Hum. Percept. Perform. **16**, 227–247.

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

't Hart, J., Collier R., and Cohen A. (**1990**). *A Perceptual Study of Intonation—An Experimental-Phonetic Approach to Speech Melody* (Cambridge University Press, Cambridge).

Xu, Y. (**1994**). ''Production and perception of coarticulated tones,'' J. Acoust. Soc. Am. **95**, 2240–2253.

Xu, Y. (**1997**). ''Contextual tonal variations in Mandarin,'' J. Phonetics **25**, 61–83.

Xu, Y. (**1998**). ''Consistency of tone-syllable alignment across different syllable structures and speaking rates,'' Phonetica **55**, 179–203.

Xu, Y. (**1999**). ''Effects of tone and focus on the formation and alignment of F_0 contours,'' J. Phonetics **27**, 55–105.

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

Xu, Y. (in press). ''Sources of tonal variations in connected speech,'' J. Chin. Ling.

Xu, Y., and Sun, X. (**2000**). ''How fast can we really change pitch? Maximum speed of pitch change revisited,'' Proceedings of The 6th International Conference on Spoken Language Processing, Beijing, pp. 666–669.

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