Proceedings of The 6th International Conference on Spoken Language Processing, Beijing. III-666-669.

HOW FAST CAN WE REALLY CHANGE PITCH? MAXIMUM SPEED OF PITCH CHANGE REVISITED

Yi Xu and Xuejing Sun

Department of Communication Sciences and Disorders, Northwestern University, USA
xuyi@northwestern.edu

ABSTRACT

The present paper reports preliminary data obtained in a study of maximum speed of pitch change. The study used an imitation paradigm to elicit fast alternating high and low pitch sequences from native speakers of Mandarin and English who were not professional singers. The speed of pitch change was measured both in terms of response time — time needed to complete the middle 75% of a pitch shift, as defined in previous studies, and in terms of excursion time — time needed to complete the entire pitch shift. Results show that the latter is nearly twice as long as the former, indicating that the maximum speed of pitch change is not nearly as fast as previous data may have implied. Potential implications of this finding on our understanding of F_0 contour production in speech are discussed.

1. INTRODUCTION

The maximum speed of pitch change was investigated in two studies by Ohala, and Ewan [1] and by Sundberg [2]. Both studies estimated the maximum speed of pitch change by measuring "response time" — time used by subjects to complete the fastest portion (the middle 75%) of a pitch shift, as shown in Figure 1. While the estimates obtained in these studies establish some important facts, certain critical aspects about the speed of pitch change are missing. First, as can be seen in Figure 1, response time does not fully reflect the fastest pitch movements 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 a speaker to complete 100% of a pitch shift, which is potentially much longer, as Figure 1 clearly indicates.

The time needed to fully complete a pitch shift is potentially very important for our understanding of pitch production in speech. In tone languages, for example, utterances often contain alternate high and low pitch sequences. As has been observed, the transitions between high and low pitches take considerable amount of time, and the duration of the transition apparently affects the shape of F_0 contours [4, 6]. Furthermore, the minimum duration of the pitch transition may also limit how close adjacent F_0 peaks and valleys can be from each other in time. Such limit may play a role in determining the alignment of F_0 peaks and valleys relative to segmental units such as syllables [5, 6, 7].

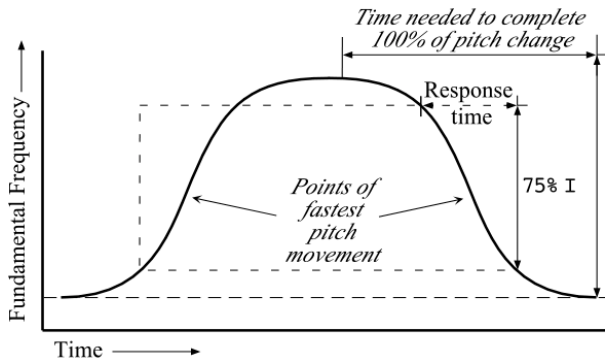

Figure 1: Measurements of speed of pitch change by Ohala and Ewan (1973) and Sundberg (1979), and by the present study (in italic).

There is probably a reason why Ohala and Ewan [1] and Sundberg [2] did not attempt to measure the time interval of complete pitch changes. In Figure 1, at the beginning of a pitch shift the F_0 movement seems to continually increase in speed, while near the end of the shift the speed of F_0 movement slows down gradually as the target register is being reached. To measure the time needed to complete 100% of a pitch shift, one has to determine the exact points in time when the shift begins and when it ends. However, the asymptotes near the onset and offset of the shift, as can be seen in Figure 1, make it difficult to locate the exact starting and ending points.

To reduce the uncertainty in determining the end points of a pitch shift, we need to minimize the duration of these asymptotes. One way to do that is to make the speaker produce a quick succession of high and low pitch registers. In this way the lingering time on each pitch register may be reduced to a minimum, resulting in a pitch contour that repeatedly goes up and down. We may refer to the F_0 patterns produced by such maneuvers as *pitch undulation*. In the present study, an experiment was conducted using an imitation paradigm to elicit pitch undulation patterns from which maximum speed of pitch changes was measured.

2. METHODS

2.1. Stimuli

The stimuli were model pitch undulation patterns based on resynthesized human voice samples. The voice samples were sustained schwas produced by a male and a female speaker. The resynthesis was done using the Praat program[1]. Based on the voice samples, vowels with various steady state fundamental frequencies were generated. For each vowel at a particular fundamental frequency, twelve model pitch variation patterns were generated. These pitch patterns differed in pitch variation pattern: /HLHLH/ or /LHLHL/, where H and L represent relatively high and low pitch registers, in pitch variation interval: 4, 7, or 12 semitones, and in undulation frequency (i.e., the number of HL or LH

[1] 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.

cycles per second): 4 or 6 Hz.

2.2. Subjects

Nineteen native speakers of American English and twenty native speakers of Mandarin Chinese between the age of 18 and 36 recruited from Northwestern University campus participated in the experiment. Speakers of the two languages were used in order to examine possible influence of language background on the speed of pitch change. Mandarin has lexical tones that use pitch patterns to differentiate words, while English only has pitch accents related to word stress. Both sexes were included in the experiment in order to examine potential gender effect. While some of the subjects had musical or voice training of some kind, none of them were professional singers or involved in a professional singing group. The tasks of the experiment turned out to be too difficult for a few subjects. As a result, only 34 subjects generated data suitable for analysis. Of the remaining subjects, 16 are English speakers (8 females and 8 males) and 18 are Chinese speakers (11 females and 7 males).

2.3. Procedure

The experiment was conducted in the Speech Acoustics Laboratory at Northwestern University. The subject was seated in a sound-treated booth in front of a computer monitor. A condenser microphone was used for the recording, and the vocalization was digitized using the SoundEdit program (Macromedia Inc.) and stored in AIFF format on a Macintosh G4 computer.

The experiment procedure was controlled by a set of HTML file, which were displayed by Netscape Navigator on a separate Macintosh G4 computer. 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. The experimental stimuli were organized into three HTML pages, each containing model undulation patterns with the pitch interval 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, and as accurately as possible in terms of both pitch interval and undulation frequency.

To examine the possibility that syllable structure may hinder or facilitate the production of pitch undulation, subjects were asked to imitate the undulation patterns both with a sustained schwa and with the syllable sequence /malamalama/.

2.4. F_0 Extraction and Measurement

The F_0 extraction was done using a procedure similar to the ones used in previous studies by the first author [4-7]. After the F_0 curves were extracted, a set of custom-written MatLab procedures were used to take the following measurements, most of which are illustrated in Figure 2.

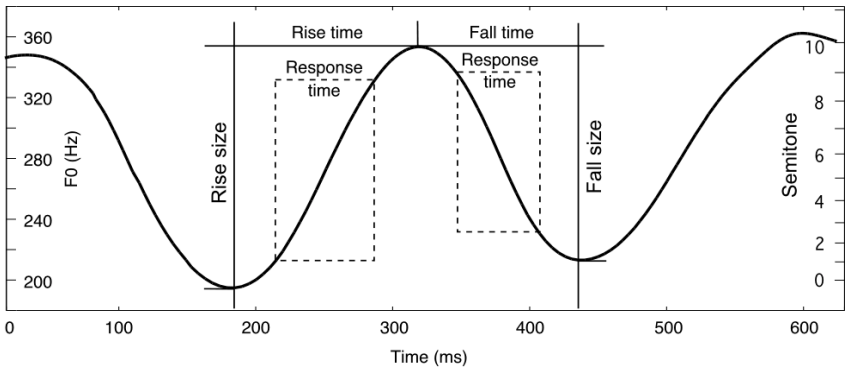

Figure 2: Measurement of excursion time and response time.

excursion size (rise or *fall)* — pitch difference (in st) between adjacent F_0 minimum and maximum in the middle undulation cycle.

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

response time: time interval corresponding to middle 75% of *excursion size* [cf. 1, 2].

excursion speed = excursion size / excursion time.

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 derivatives of the F_0 curves.

3. ANALYSES AND RESULTS

Reported in the following are the results of preliminary analyses performed on the measurements listed above.

Table 1 displays the means of excursion size, excursion time, excursion speed, response time, and maximum velocity broken up according to language, gender, carrier, (excursion) direction, and (pitch-shift) interval. 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 1 it can be seen that the effect of interval is significant for all the measurements. This indicates not only that subjects managed to produce different excursion sizes for the three pitch-shift intervals, but also that the speed of pitch change varied significantly across the intervals. The effect of language is significant for four of the measurements. Note that excursion time is longer for English subjects than for Mandarin subjects. This may at first lead us to believe that English speakers are slower in pitch change. However, their speed of pitch change is actually faster as indicated both by excursion speed and maximum velocity. It appears that the larger excursion size of the English subject actually gives rise to faster speed. This possibility is verified by the correlation coefficients shown in Table 2, where excursion size is highly correlated with both

	language		**gender**		**direction**		**carrier**		**interval**		
	Chinese	English	female	male	rise	fall	schwa	mala	4	7	12
excursion size (st)	4.3	5.8	5.1	4.9	4.8	5.3	5.2	4.9	4.0	4.7	6.4
	p = .0001		p = .3462		p < .0001		p = .0448		p < .0001		
excursion time (ms)	125.0	138.7	128.5	135.2	131.0	131.8	132.5	130.3	126.4	128.8	139.1
	p = .0112		p = .291		p = 0.6039		p = .3906		p < .0001		
excursion speed (st/s)	35.6	42.7	40.8	36.6	36.9	40.9	39.1	38.8	31.9	37.2	47.7
	p = .0363		p = .177		p < .0001		p = .465		p < .0001		
response time (ms)	65.7	70.4	67.7	68.2	67.1	68.8	68.7	67.1	65.2	67.0	71.5
	p = .0725		p = .9762		p = 0.1562		p = .328		p < .0001		
max velocity (st/s)	63.4	76.5	73.6	64.4	64.8	74.3	70.3	68.8	56.1	65.4	87.1
	p = .0203		p = .0898		p < .0001		p = .3535		p < .0001		

Table 1: Mean values of various measures under the effects of language, gender, (excursion) direction, carrier, and pitch shift interval, together with probability values resulting from five-factor mixed-measure anovas.

excursion speed and maximum velocity, but excursion time is correlated with neither of them, despite the fact that time is actually in the equation for computing excursion speed. 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.[2]

	size	time	speed	velocity	response
(excursion) size	1.000	.337**	.905**	.889**	.251**
(excursion) time		1.000	-.054	.027	.939**
(excursion) speed			1.000	.937**	-.109*
(max.) velocity				1.000	-.059
response (time)					1.000

Table 2: Correlation of various factors (computed from 408 observations). ** means p < 0.01 and * means p < 0.05.

The effect of direction is significant for three of the measurements in Table 1. While the fact that falls are faster than rises is consistent with the findings of Sunberg [3], like the language difference, it is also possibly related to the relatively larger excursion size of falls. The effect of carrier is only marginally significant for one of the measurements. So, at least for the present experiment, whether the carrier was a sustained schwa or a syllable sequence did not make much difference. Surprisingly different from [3], the effect of gender was not significant for any of the measurements.

direction	rise			fall		
interval	4	7	12	4	7	12
excursion size	3.7	4.5	6.3	4.2	5.0	6.6
excursion time	125	127	141	127	131	138
response time	64	65	72	67	69	71
excursion / response	1.95	1.95	1.96	1.90	1.90	1.94

Table 3: Excursion time, response time, and ratio of excursion time to response time.

Perhaps the most interesting results come from the comparison between response time and excursion time. Table 3 displays the mean values of response time and excursion time broken up by direction and interval, and the ratio of excursion time to response time. Recall that excursion time is the amount of time used to complete 100% of a pitch shift, while response time is the amount of time used to complete the fastest middle 75% of a pitch shift. As shown in Table 3, for all excursion sizes, excursion time is nearly twice as long as response time. This means that the onset and offset of a pitch shift, which cover only 25% of the entire pitch change, take almost as much time as it takes to complete the center 75% of the pitch shift.

4. DISCUSSION

The results of the preliminary data analyses indicate that a pitch shift involves not only a fast F_0 movement, but also an onset and an offset which together cover only a small portion of the pitch interval but take about just as much time. This finding may seem surprising at first. But it makes sense if we consider what probably has to happen when a rapid but accurate pitch shift is produced. First the larynx needs to accelerate from the starting pitch level to reach the maximum speed of pitch change. Then, as the target pitch level is being approached, the larynx needs to decelerate in order not to overshoot the target.

The finding that the onset and offset of a pitch change take almost as long as response time is significant. It means that if the response time values reported by earlier studies [1,2] were taken as a direct indicator of maximum speed of pitch change, the time needed for accomplishing an entire pitch shift would have been substantially underestimated. This has indeed been the case with 't Hart, J., R. Collier, and Cohen [3]. They argue in their book that it is perception rather than production mechanisms that are responsible for the observed maximum rate of pitch change in speech (pp. 71-75). The argument was based on the authors' interpretation of the data on speed of pitch change reported by Sundberg [2]. They used the response time for the 12 st condition to compute the maximum speed of pitch change, and arrived at the value of 120 st/s. Because the fastest pitch movement they observed in Dutch was 50 st/s, they concluded that articulatory limits simply could not have been responsible for the observed rate of pitch change in speech. Instead, they argued, 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. Unfortunately, their use of the response time values may have exaggerated the maximum speed of pitch change in two ways. First, as found by all three studies on speed of

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

pitch change [1, 2, present], the rate of pitch change is faster for a larger pitch interval than for a smaller one. So, unless a particular pitch movement actually covers 12 st, it is inappropriate to use the data for that interval as the indicator of speed of pitch change at other intervals. Second, as found in the present study, the full excursion time is about twice as long as response time. Thus a more accurate formula for converting response time to excursion speed should be

excursion speed = interval / (1.93 • response time)

where 1.93 is the mean ratio of excursion time to response time shown in Table 3. Using this formula, estimated pitch excursion speed are computed from Sundberg's data and compared with the excursion speed obtained in the present study, as shown in Table 4.

direction	rise			fall		
interval	4	7	12	4	7	12
present study	29.9	35.5	45.4	33.8	38.9	50.0
Sundberg 1979	25.0	40.8	59.8	28.4	50.4	82.9

Table 4: (a) excursion speed (st/s) obtained in present study for pitch rises and falls in three interval conditions. (b) excursion speed computed from [2] using the formula: *excursion speed = interval / (1.93 • response time).*

It may appears at first that the excursion speeds in our study are slower than those of Sundberg's. But this is not really the case. In Table 4 the excursion speeds for the present study are computed with the actual excursion sizes achieved by the subjects as shown in Table 3, not the required intervals. Thus our 12-st condition should parallel Sundberg's 7-st condition. Taking this into consideration, the excursion speeds from the two studies are fairly comparable. Furthermore, because the full excursion size found in Dutch by 't Hart et al. is around 6 st [3] (p. 53), in general, therefore, the speed of pitch change in Dutch is also comparable to the excursion speeds in Table 4. Just as interestingly, 't Hart et al. also report that in English, full-size rises and falls can cover an octave and the rate of change is about 75 st/s (p. 49). This, again, is comparable to the excursion speed computed from Sundberg's data for the 12-st condition as shown in Table 4.

As for whether maximum speed of pitch may play a role in determining the shape and alignment of F_0 contours in speech, there have already been evidence that it could be the case. The results of the present study seems to be able to provide more concrete links between the physical limits as indicated by the maximum speed and certain F_0 variations in speech. A case in point is the seemingly long carryover F_0 transitions found in Mandarin in recent studies [4, 6]. For example, in a L-H tone sequence, the initial portion of the F_0 contour corresponding to H is usually a long rising ramp. As indicated by Table 3, it takes about 125-141 ms to raise pitch by 3.6-6.3 st. This means that an apparent rising transition is simply inevitable in a L-H sequence which usually covers a pitch range of about 6 st [6]. It further suggests that similar F_0 transitions should also occur in any utterance in any language that contains alternating high and low pitch targets.

Another case is the phenomenon of peak delay in Mandarin. Unlike the LRL sequence, the F_0 peak in a LHL sequence usually occurs within the H-carrying syllable. However, It was recently found that at fast speech rate, the peak may occur after the H-carrying syllable [7]. It was further found that the minimum distance between the onset of the rise and the syllable offset beyond which such peak delay would occur was 125.7 ms [7]. This distance is about the same as or shorter than the excursion times shown in Table 3. This suggests that there may be a direct link between peak delay and the maximum speed of pitch change.

5. CONCLUSIONS

The pitch undulation imitation paradigm used in the present study allowed us to measure both response time — time needed to complete the middle 75% of a pitch shift, and excursion time — time needed to complete the entire pitch shift, produced by Mandarin and English speakers. Preliminary data analyses revealed that while the response time values obtained in the present study are comparable to those of previous studies [1, 2], excursion time is roughly twice as long. This finding indicates that there is much greater physiological limitation on the maximum speed of pitch change than has been recognized. It further suggests that the role of physiological constraints in determining the shape and alignment of F_0 contours in speech is likely to be much more important than has been appreciated. Meanwhile, however, although apparent links can be already seen between the data reported here and certain phenomena of F_0 contour variations observed in real speech, more detailed analyses need to be completed before we can be more certain about the full implications of the new finding on our understanding of F_0 contours in speech.

6. ACKNOWLEGEMENT

This work is supported by NIH Grant DC03902.

7. REFERENCES

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

2. Sundberg, J. "Maximum speed of pitch changes in singers and untrained subjects," *J. Phon*. 7, 71-79, 1979.

3. 't Hart, J., Collier, R. and Cohen, A. *A perceptual Study of Intonation — An experimental-phonetic approach to speech melody*, Cambridge University Press, Cambridge, 1990.

4. Xu, Y. "Contextual tonal variations in Mandarin," *J. Phon*. 25, 61-83, 1997.

5. Xu, Y. "Consistency of tone-syllable alignment across different syllable structures and speaking rates," *Phonetica* 55, 179-203, 1998.

6. Xu, Y. "Effects of tone and focus on the formation and alignment of F_0 contours," *J. Phon*. 27, 55-105, 1999.

7. Xu, Y. "Fundamental frequency peak delay in Mandarin," to appear in *Phonetica*.