# A Perception Study on the Third Tone in Mandarin Chinese

Rui Cao
*Doctoral Student, Ph.D. Linguistics*
*University of Florida*

Priyankoo Sarmah[1]
*Doctoral Candidate, Ph.D. Linguistics*
*University of Florida*

**Abstract:** This experimental study examines the role of the shape of the pitch contour in the perception of the Mandarin Chinese tone 3.[2] A set of stimuli was constructed by varying the pitch of tone 3 on two conditions: (1) varying the duration of the dip (or turning point) and (2) varying the timing of the turning point (duration of the slope). The manipulated pitch contours of tone 3 were presented to the native speakers of Mandarin Chinese in two sets: (a) a set of speech stimuli and (b) a set of non-speech stimuli. The participants of the experiment were asked to perform a judgment task in order to identify the tone. The results were analyzed and it was found that there is a specific range of the stimuli in both conditions where tone 3 is perceived by native Mandarin Chinese speakers.

## 1.0 Introduction

Perception of lexical tones by both tone and non-tone language speakers has been of interest in the domain of auditory phonetics. Among tone languages, Mandarin Chinese has received considerable interest from auditory phoneticians due to its complex and varied tonal inventory.

In Mandarin Chinese there are four lexical tones varying in contour shapes: a high-level tone (tone 1, 55), a mid, rising tone (tone 2, 35), a low-falling-rising tone (tone 3,

[1] The authors would like to thank Dr. Ratree P Wayland and Dr. Caroline R Wiltshire of the Program in Linguistics, University of Florida for their comments and support, the Chinese students at University of Florida who participated in this study, the attendees at UTASCIL-14 for their comments and suggestions, and the sponsor of the Yumi Nakamura award. Financial assistance received from the award was used to compensate the participants in this study. The authors are indebted to the two anonymous reviewers for their comments and suggestions.

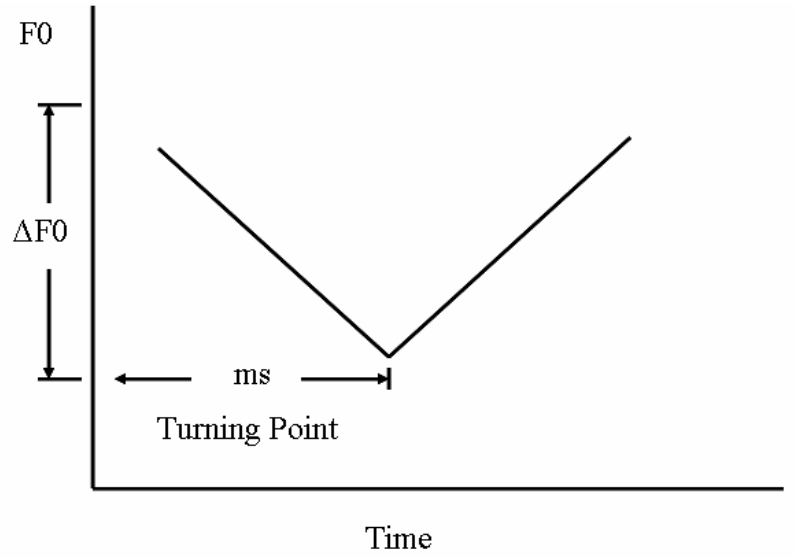
[2] In this paper the terms *tone 1*, *tone 2*, *tone 3* and *tone 4* will be used consistently to indicate the high-level, rising, low-falling-rising and high-falling tones respectively, in Mandarin Chinese.

214), and a high-falling tone (tone 4, 51). However, Mandarin Chinese tone shapes in natural sentences often deviate from their expected canonical shapes, motivating investigations into their perception by native speakers of the language (Shih and Kochanski, 2000). Shen and Lin (1991) have reported that besides the F0 height, which contributes to the distinguishing of these two tones, there are two other parameters that are relevant: (i) the timing of the turning point, defined as the duration from the onset of the tone to the point of change in F0 direction, and (ii) the decrease in F0 from the onset of the tone to the turning point, which they call the ΔF0 (see Figure 1).

Shen and Lin (1991) found that when ΔF0 is set at 30 Hz, native speakers of Mandarin Chinese perceive tone 3 when the turning point is more than 40% of the total length of the stimuli. Again, when ΔF0 is 15 Hz, tone 3 is perceived when the turning point is more than 60-70% of the total length of the stimuli. However, Shen and Lin (1991) were not specific about the geographical affiliation of the Mandarin Chinese speakers who participated in their study. Marked idiosyncrasies among Mandarin Chinese speakers in producing tone 3 can be safely considered as areal features; e.g., Yip (2002) reports that the Tianjin variety of Mandarin Chinese differs significantly from the Beijing variety[3] in terms of the production of tone 3.

[3] In several personal communications with native speakers of Mandarin Chinese, it has been pointed out to these authors that tone 3 of Mandarin Chinese produced by northern Chinese speakers is noticeably different from the one produced by southern Chinese speakers. Unlike the falling-rising contour seen among the northern Mandarin Chinese speakers, southern speakers produce a low-falling tone for tone 3 of Mandarin Chinese.

*Figure 1: Turning point and ΔF0 properties schematized for a contour tone*




Moore and Jongman (1997) report a simultaneous effect of timing of the turning point and ∆F0 in Mandarin Chinese speakers' perception of tone 2 and tone 3. In one of their experiments, they manipulated the timing of the turning point and ΔF0 systematically in isolated synthetic speech stimuli. They first created 12 stimuli having turning points at various times from 20 to 240 ms in 20 ms steps, and then each stimuli was varied in terms of ∆F0 ranging from 10 to 70 Hz in 5 Hz steps. All the 156 stimuli (12 X 13) were presented to native speakers of Mandarin Chinese in a forced choice perception test where lexical entries with tone 2 (无 'not') and tone 3 (舞'dance') were the choices.

The results showed that when the turning point is less than 240 ms and ∆F0 is below 30 Hz, more tone 2 is perceived. When the turning point is more than 200 ms and ∆F0 is higher than 35 Hz, subjects record tone 3 responses. They conclude that ∆F0 becomes crucial in the perception of tone 3 when the turning point is late; however, in that case ∆F0 needs to be more than 35 Hz.

It has also been reported that phonetic similarities between tone 2 and tone 3 in Mandarin

Chinese make categorical distinction difficult in both first language and second language acquisition (Liu, 2004).

The aforementioned studies prompt one to understand the perception of Mandarin Chinese tones in further detail. In the present study, the perception of tone 3 of Mandarin Chinese by its native speakers is studied. It is expected that this study will be able to show that (a) duration of the turning point and (b) the timing of the turning point, both are important for accurate identification of tone 3 in Mandarin Chinese.

## 2.0 Methodology

A set of 80 speech stimuli was constructed with varying pitch contours on the syllable [ma] (see section 2.1.1). Another set of 80 non-speech stimuli was constructed, devoid of any consonantal or vocalic information but with only pitch contours on them (see section 2.2.2). The two sets of stimuli were administered separately to 9 native Mandarin Chinese speakers from the Beijing area who were asked to perform a tone identification task. Each stimulus was presented 10 times and all the stimuli in each set were randomized.

### *2.1 Subjects*

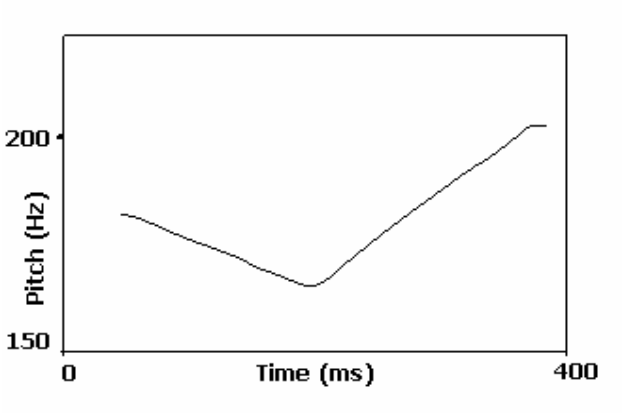
Nine subjects, four females and five males, between the ages of 22 and 35, participated in the study. All of them were from Beijing, China, and had studied at the University of Florida as graduate students for 6 months to 5 years. Mandarin Chinese is the only language, besides English, that they speak. None reported speech or hearing disorders. Four of the subjects were paid for their participation and five participated voluntarily.

### *2.2 Stimuli*

#### *2.2.1 Experiment 1*

An isolated utterance of [ma] (tone 3, 马 'horse') from a female native speaker of Mandarin Chinese from the Beijing area was recorded, using the PRAAT software (Paul Boersma and David Weenink) at a sampling frequency of 44100 Hz. The total length of the sound was 400 ms with a turning point at 200 ms. In the original iteration of the "ma", three pitch heights were observed: onset F0 = 186Hz, offset F0 = 211 Hz and the turning point F0 = 163 Hz (see Figure 2). Without manipulating the length, F0 on the onset and offset of the recorded utterance, two parameters were altered to create two sets of stimuli: the duration of the dip (condition 1) and the timing of the turning point (condition 2).

*Figure 2: Pitch Track of the Original Tone 3*

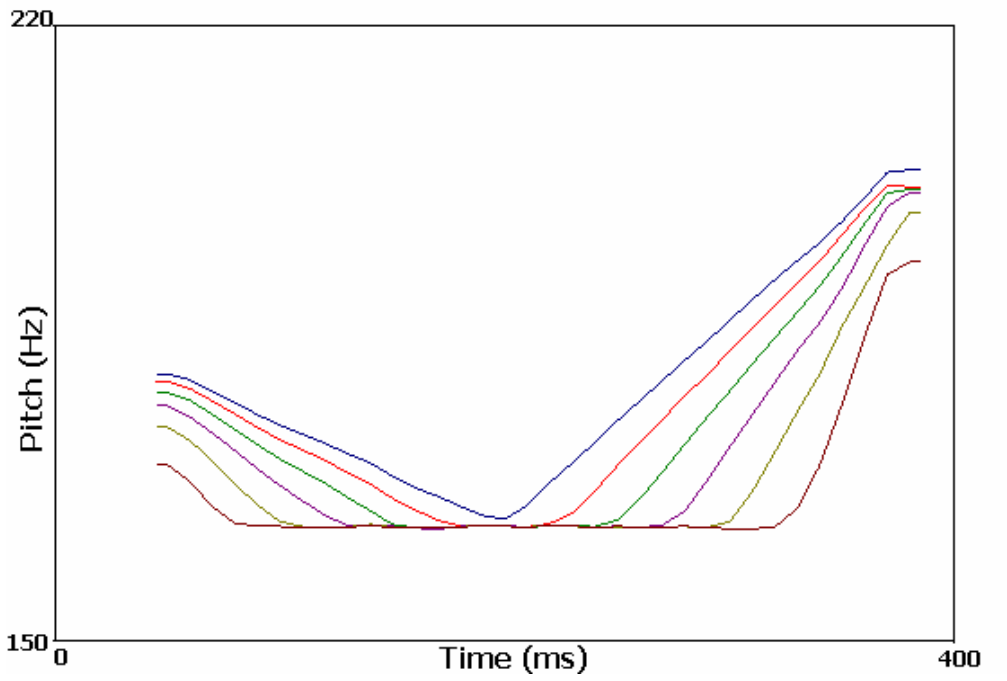



For the stimuli in condition 1, the duration of the dip was increased from the midpoint of the utterance by an increment of 10 ms (5 ms on both sides) until it was equal to the total duration of the stimuli. This resulted in 40 stimuli having dip duration from 0 ms to 390 ms. Table 1 shows a sample of the points added in different timings in the empty pitch tier for the stimulus. Figure 3 provides graphical representations of some of the pitch contours after manipulation.

*Table 1: Points added for generating stimuli*

| Stimuli | Starting Point of the Dip | Ending Point of the Dip | Duration of the Dip |
|---|---|---|---|
| 1 | 195 ms | 205 ms | 10ms |
| 2 | 190 ms | 210 ms | 20 ms |
| 3 | 185 ms | 215 ms | 30 ms |

*Figure 3: The Duration of the Dip Altered*

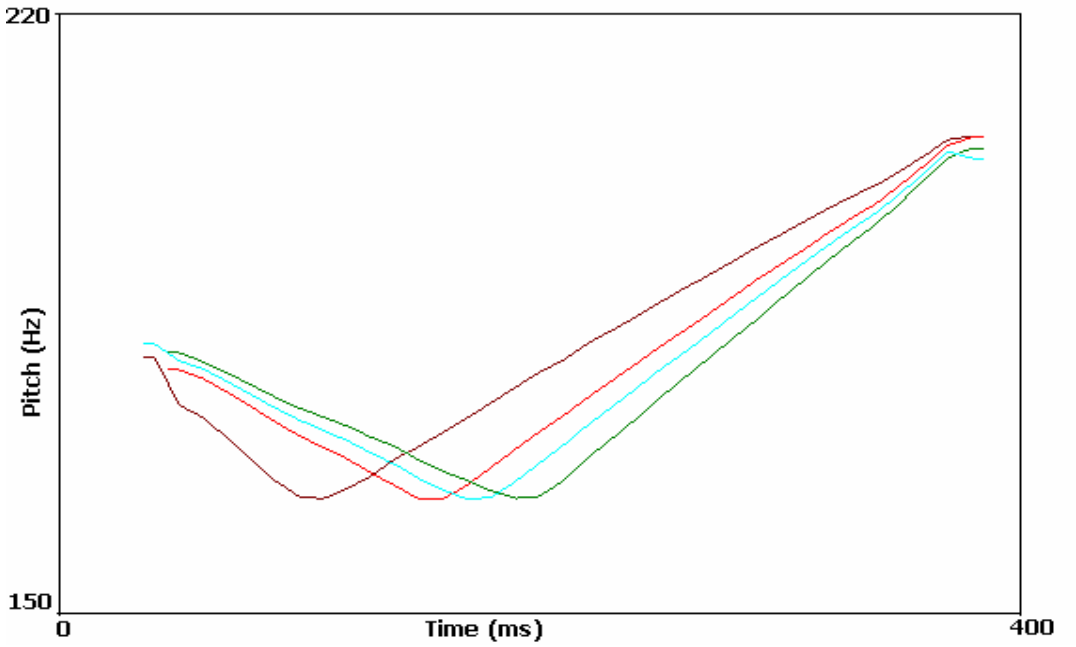



For the stimuli in condition 2, the dip was kept constant at 10 ms; only the timing of the turning point was altered. For the down slope, the duration was altered from 0 to 390 ms with an increment of 10 ms resulting in 40 stimuli in total. Table 2 shows a sample set of the 40 stimuli. Figure 4 provides a graphical representation of the manipulated pitch contours where the timing of the turning point is altered.

*Table 2: Points added for generating stimuli*

| Stimuli | Onset | Starting Point of the Dip | Ending Point of the Dip | Offset | Duration of the Down slope |
|---|---|---|---|---|---|
| 1 | 186 Hz | 0 ms | 10 ms | 211 Hz | 10 ms |
| 2 | 186 Hz | 10 ms | 20 ms | 211 Hz | 10 ms |
| 3 | 186 Hz | 20 ms | 30 ms | 211 Hz | 10 ms |

*Figure 4: The Turning Points (Duration of the Slopes) Altered*




*2.2.2 Experiment 2*

The 80 stimuli in experiment 1 were used for experiment 2. However, in experiment 2, there was no consonantal or vocalic information provided in the stimuli. Keeping the pitch contours identical to the stimuli in the two conditions in experiment 1, sine waves were generated devoid of any consonantal or vocalic information. This was achieved by manipulating the stimuli used in experiment 1 with the help of the PRAAT software.

### *2.3 Procedure*

After preparing the two sets of stimuli, PRAAT was used to present the stimuli to the subjects and record their responses in this study. Four participants participated in both experiment 1 and experiment 2, whereas the other five participated only in experiment 1. To avoid contiguity effects, each stimulus was repeated 10 times and randomly played on a computer using PRAAT. Participants listened to the stimuli through a pair of headphones. There were breaks after every 250 tokens during the experiment to make sure that the participants were not fatigued by continuously listening to the stimuli, resulting in false choices by them.

After each token, a subject would give a response of what tone s/he heard by clicking one of the boxes on the computer screen, labeled "first", "second", "third", "fourth' or "not sure".[4] In a pre-test it was ascertained that the participants could correctly associate the four Mandarin Chinese tones with the numerical representation system of the tones of the language.

Each experiment lasted for about 20 minutes. After a participant completed her/his experiment, the results were saved on the hard disc of a computer in a tabulated form for further analysis.

## 3.0 Results

### *3.1 Experiment 1*

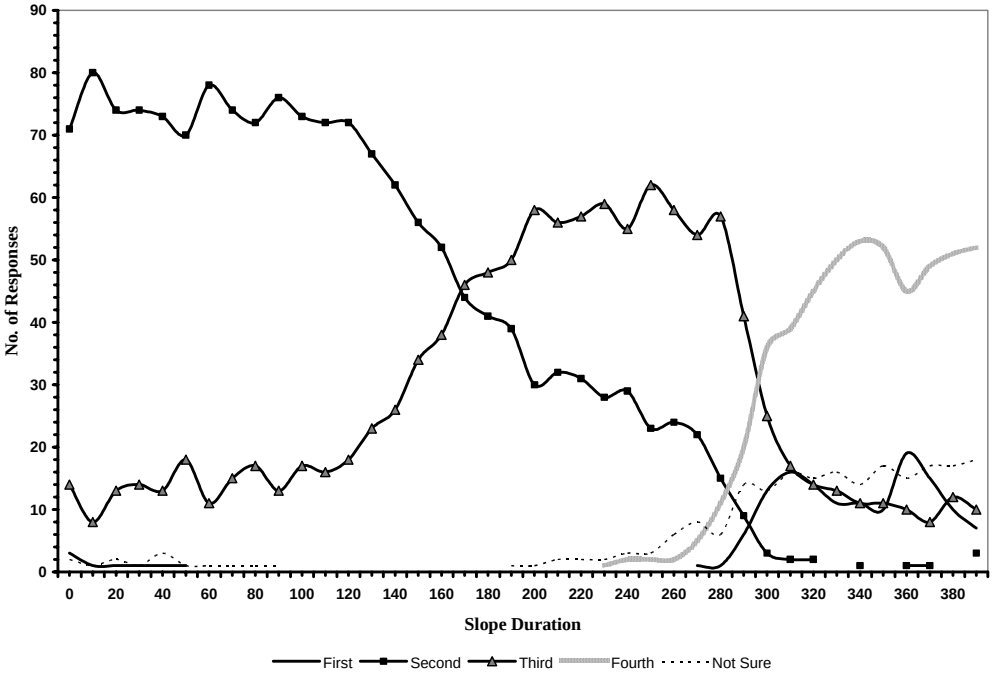
For experiment 1 condition 1 (see Figure 5), it was observed that the stimuli with dip duration between 0 ms and 270 ms, were mostly perceived as a tone 3; when the duration of

[4] As this study is aimed at investigating Mandarin Chinese speakers' perception of tone 3, the participants are not asked to further extend the association of the [ma] stimuli with a particular lexical meaning.

the dip is greater than 270 ms, the stimuli were mostly perceived as a tone 1.

*Figure 5: Response of Subjects for Experiment 1 Condition 1 for [ma]*


*Figure 6: Response of Subjects for Experiment 1 Condition 2 for [ma]*

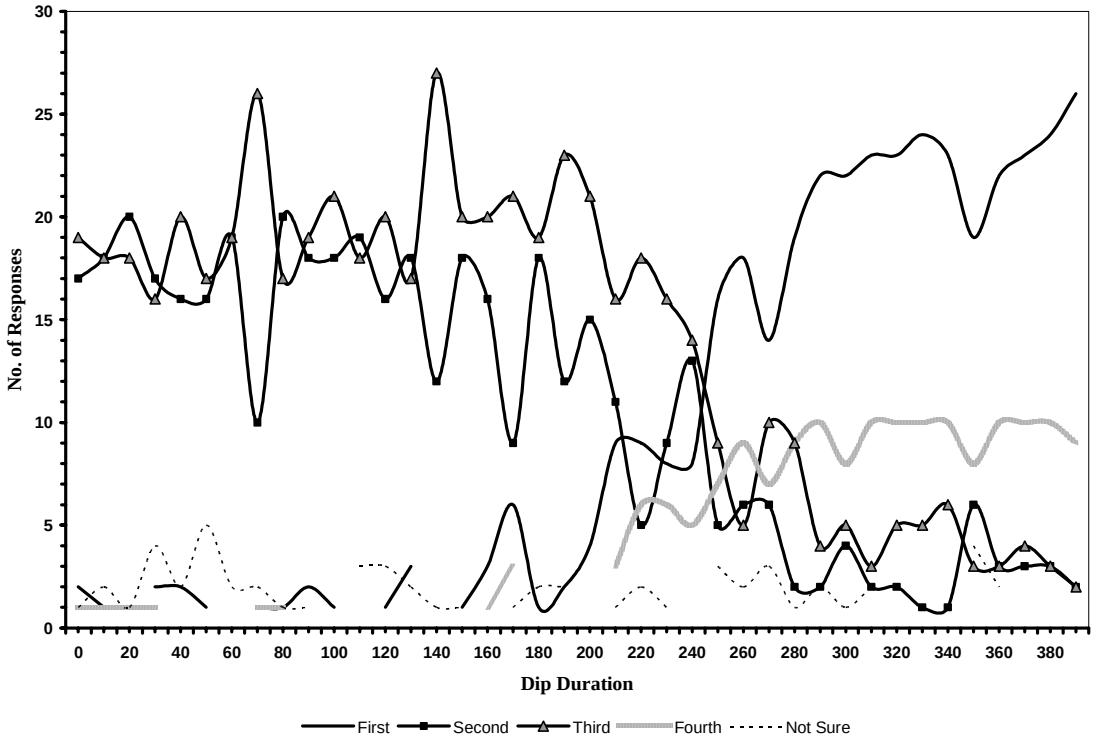


In condition 2, when the timing of the turning point varied from 170 ms to 290 ms, the stimuli were mostly perceived as tone 3; when the turning point was less than 170 ms, tone 2 was perceived, and when the turning point occurred after 290 ms, the stimuli were perceived as a tone 4 of Mandarin Chinese.

### *3.2 Experiment 2*

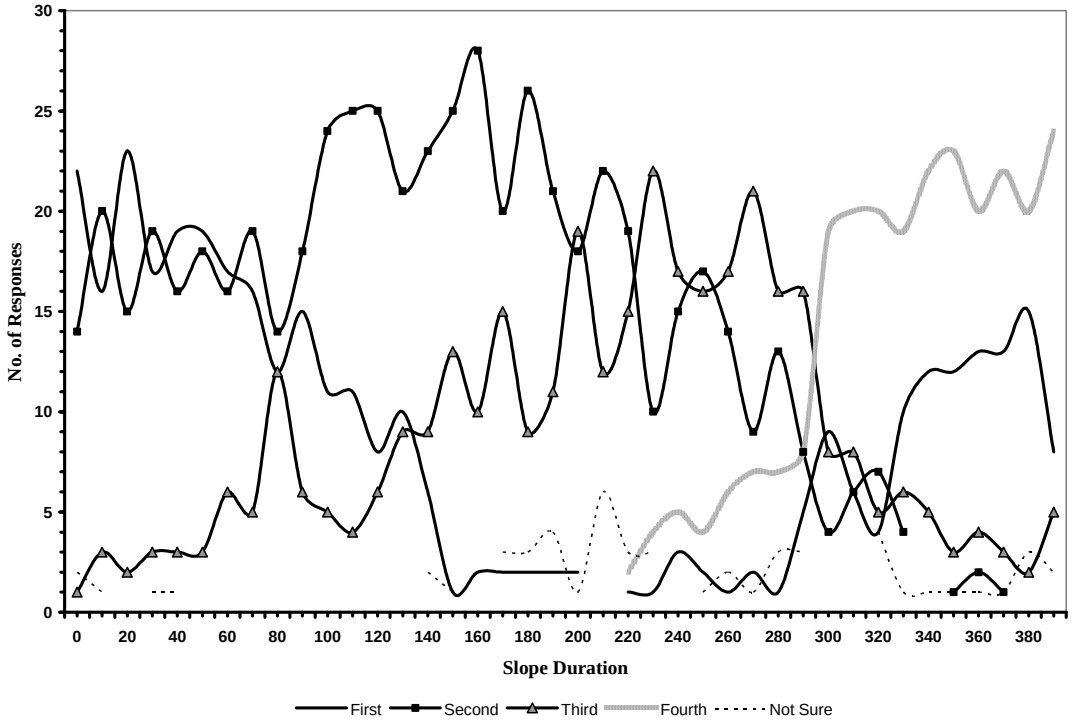
In Experiment 2, four participants listened to pure tones with the same pitch contours as in experiment 1. In this experiment, for condition 1, the participants consistently demonstrated mixed perception of tone 2 and tone 3 until the dip was 250 ms long. Though the total number of responses was more for tone 3 than for tone 2, no apparent consistency was demonstrated in favor of tone 3 (see Figure 7). However, if the slope is longer than 250 ms, the participants consistently perceived a tone 1 of Mandarin Chinese.

*Figure 7: Response of Subjects for Experiment 2 Condition 1 for sine tones*



In condition 2, the participants demonstrated a mixed perception of tone 1 and tone 2 until the timing of the turning point was 90 ms (Figure 8). However, when the turning point was between 90 ms and 220 ms, more tone 2 were perceived. Between 230 ms and 290 ms, more tone 3 was perceived. Turning point timing more than 290 ms evoked a tone 4 perception among the participants.

*Figure 8: Response of Subjects for Experiment 2 Condition 2 for sine tones*




### *3.3 Summary of responses*

*3.3.1 Experiment 1: [ma] syllable stimuli*

*Table 3: Experiment 1, Condition 1*

| **Condition 1** | |
|---|---|
| Dip duration | Perceived as |
| 0-270 ms | Tone 3 |
| 280-390 ms | Tone 1 |

*Table 4: Experiment 1, Condition 2*

| **Condition 2** | |
|---|---|
| Timing of the turning point | Perceived as |
| 0-160 ms | Tone 2 |
| 170-290 ms | Tone 3 |
| 300-390 ms | Tone 4 |

*3.3.2 Experiment 2: Non-speech (sine wave) stimuli*

*Table 5: Experiment 2, Condition 1*

| **Condition 1** | |
|---|---|
| Dip duration | Perceived as |
| 0-250 ms | Mixed (Tone 2 and Tone3) |
| 260-390 ms | Tone 1 |

*Table 6: Experiment 2, Condition 2*

| **Condition 2** | |
|---|---|
| Timing of turning point | Perceived as |
| 0- 80 ms | Mixed (Tone 1 and Tone 2) |
| 90-220 ms | Tone 2 |
| 230-290 ms | Tone 3 |
| 300-390 ms | Tone 4 |

## 4.0 Discussion

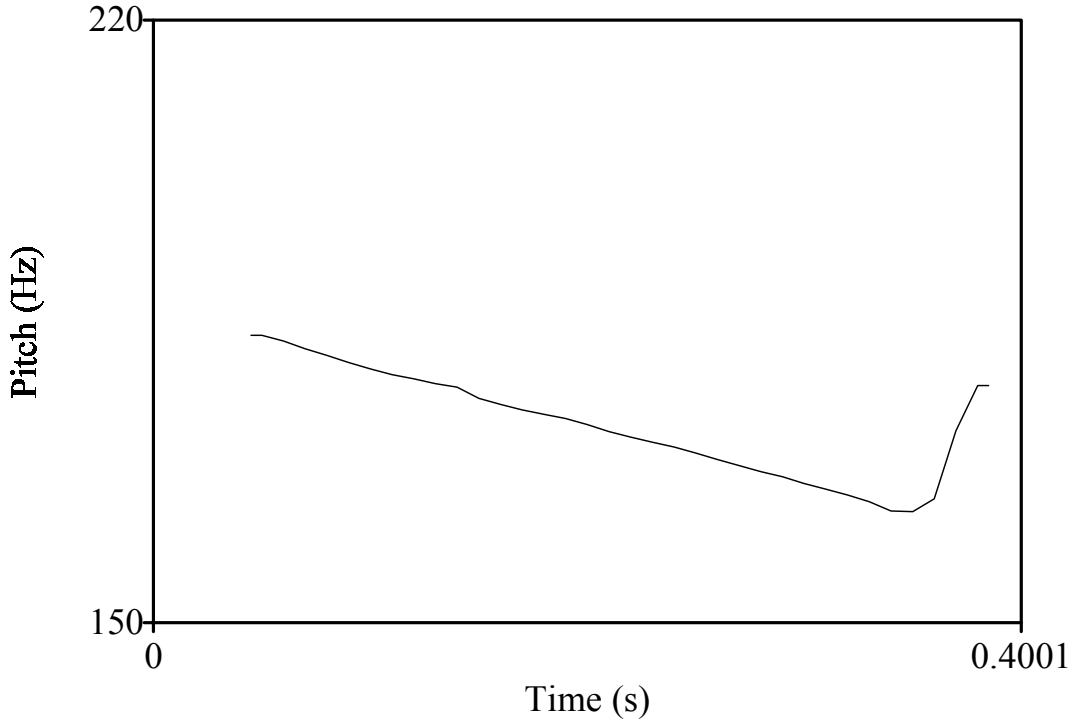
The current study examined the role of duration of the dip and the timing of the turning point in perceiving tone 3 of Mandarin Chinese. Results suggest that both factors play significant roles in the perception of the tone 3 in the language.

### *4.1 Experiment 1*

In the case of the speech stimuli [ma], the duration of the dip of the tone 3 in Mandarin Chinese should not be more than 67.5% of the total length of the stimuli. Once the dip is more than 67.5% of the total length of the stimuli, Mandarin Chinese speakers perceive the stimuli as tone 1.

In the case of the speech stimuli, a tone 3 is perceived if the turning point occurs between 42.5% and 72.5% of the total length of the stimuli. If it occurs before 42.5%, it is most likely to be perceived as a tone 2 and if it occurs after 72.5% of the total length of the stimuli, the stimuli is perceived as tone 4. It is noteworthy here that Shen and Lin (1991) did not report any perception of tone 4 by their speakers. However, as demonstrated in our results, a manipulated falling rising pitch contour may be perceived by native speakers of Mandarin Chinese as tone 4, depending on the timing of the turning point. As the turning point of the falling-rising contour gets delayed (e.g. Figure 9), perceptually it resembles the pitch contour of tone 4 in Mandarin Chinese. This explains the tone 4 responses recorded by the participants of this study.

*Figure 9: A Pitch Contour with a late turning point*




The absence of tone 4 judgments in Shen and Lin (1991) can be attributed to the binary

forced-choice identification test they adopted in their study where participants were forced to make a choice between tone 2 and tone 3 responses. Under such experimental conditions it is plausible that some tone 4 perceptions were incorrectly attributed to tone 2 or tone 3 responses in their study.

### *4.2 Experiment 2*

Initially we could not come to any clear conclusion about the range in which a tone 3 is perceived in terms of the dip duration (condition 1). However, we noticed that in the 0-62.5% range, the number of tone 3 responses is more (N=489) than the number of tone 2 responses (N=385). Moreover, a paired t-test conducted on the two sets of data (tone 2 and tone 3) confirmed that there is a significant difference ($p<0.05$) between the two sets. Hence, statistically speaking when the dip duration is up to 62.5% of the total length of the stimuli, the non-speech stimuli are perceived significantly as tone 3.

In experiment 2 as far as the timing of the turning point is concerned (condition 2), we noticed that if the turning point occurs between 57.5% and 72.5% of the total duration of the stimulus, the tone of the stimulus is considered to be tone 3. If the turning point occurs after 72.5% of the total length of the stimuli, it is perceived as tone 4. If the turning point occurs between 20% and 55% of the total length of the stimuli, it is perceived as tone 2. Any turning point that occurs before 20% of the total length of the stimuli evokes a mixed perception of tone 1 and tone 2 of Mandarin Chinese.

It is noticed that even though perception of speech stimuli in experiment 1 demonstrates clear categorical distinction, the same is not true for the non-speech stimuli in experiment 2. Experiment 2 demonstrates a considerable amount of overlapping in perception of the

non-speech stimuli. Moreover, the ranges in which tone 3 is perceived in conditions 1 and 2, differ significantly in the two experiment conditions (speech and non-speech). Considering the fact that only a small number of participants (only four) participated in experiment 2, we would not like to single out a specific factor as being responsible for the anomalies between the results demonstrated in experiment 1 and experiment 2. However, we do have a few suggestions to account for this anomaly:

(*i*) It has been noticed that perception of speech and non-speech stimuli can be significantly different. Mody et al. (1997) reported that in the perception of speech (/ba/-/da/) and non-speech stimuli, good readers made fewer errors in identifying speech stimuli. Following Mody et al. (1997), it may be suggested that the anomaly between the results of experiment 1 and experiment 2 arises due to the type of stimuli—speech and non-speech

(*ii*) The participants' tendency towards mixed responses in experiment 2 may also have certain semantic underpinnings. As participants are not able to map the non-speech stimuli, devoid of spectral information, onto any word in their native language, they are not able to categorize the sine wave stimuli in a tonal category, resulting in mixed responses. However, we are not aware of any studies which address semantics as a factor in the perception of non-speech stimuli.

From experiment 1, it can be safely concluded that timing of the turning point and duration of the dip serve as important cues in the identification of tone 3 in Mandarin Chinese. These two cues are also important to categorically distinguish tone 2 and tone 3 in Mandarin Chinese. Occurrence of these two cues outside the perceptual range of tone 2 and tone 3 can evoke the perception of tone 1 and tone 4 among Mandarin Chinese speakers. Considering

the role of tones in distinguishing lexical meanings in Mandarin Chinese, it can be assumed that in the case of speech stimuli, native speakers would associate the speech stimuli with dip durations up to 67.5% of the total duration of the stimuli with the lexical meaning 'horse' (tone 3, 马), any stimuli with a dip duration of more than 67.5% of the total duration will be associated with the lexical meaning 'mother' (tone 1, 妈). Similarly, when the timing of the turning point is between 42.5% and 72.5% of the total length of the stimuli, the stimuli will be associated with the lexical meaning 'horse' (马). However, if the timing of the turning point is less than 42.5% of the total duration of the stimuli, the stimuli will be associated with the lexical meaning 'hemp' (tone 2, 麻). If the turning point occurs after 72.5% of the total length of the stimuli, the stimuli will be associated with the meaning 'scold' (tone 4, 骂).

The results of this study also have some tertiary findings. Blicher et al. (1990) have suggested that duration serves as a prominent cue in the identification of tone 3 in Mandarin Chinese. They noticed that tone 3 is significantly longer than tone 2 in Mandarin Chinese. However, in this study we noticed that even though the total duration of the stimuli were as much as of a tone 3 monosyllable (400 ms), it could not be a distinctive cue for the perception of tone 3 as under several conditions native speakers still made judgments favoring tone 2.

A tertiary conclusion of this study is that Mandarin Chinese speakers perceive tones in speech and non-speech stimuli in markedly different ways. Non-speech stimuli devoid of consonantal and vocalic information do pose a problem for native Mandarin Chinese speakers in the accurate identification of the tone.

**References:**

Blicher, D. L., R. Diehl, and L. B. Cohen. 1990. Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. *Journal of Phonetics*, 18: 37–49.

Liu, Y-T. 2004. The Comparative fallacy in tone perception studies. *Columbia University Working Papers in TESOL and Applied Linguistics*, 4 (1).

Mody, M., M. Studdert-Kennedy, and S. Brady. 1997. Speech perception deficits in poor readers: Auditory processing or phonological coding? *Journal of Experimental Child Psychology*, 64 (2): 199-231.

Moore, Corinne B. & A. Jongman. 1997. Speaker normalization in the perception of Mandarin Chinese. *Journal of Acoustical Society of America* 102(3), September 1997.

Shen, X., and M. Lin. 1991. A perceptual study of Mandarin Tones 2 and 3. *Language Speech* 34, 145-156.

Shih , Chilin and G. Kochanski. 2000. Chinese tone modeling with Stem-ML. Paper presented at the *International Conference on Speech and Language Processing*, (ICSLP-2000), Beijing.

Yip, M. 2002. *Tone*. Cambridge University Press.