

Mandarin Lexical Tone Recognition: The Gating Paradigm

Yuwen Lai and Jie Zhang
University of Kansas

Abstract

Research on spoken word recognition in Indo-European languages often does not incorporate prosody. In Mandarin Chinese, however, lexical prosody is used extensively and has been shown to affect word processing in previous studies. The present study uses the gating paradigm to investigate the processing of the four Mandarin tones as well as the role of the initial segment in processing. Duration-blocked gates generated from eight monosyllabic quadruplets with matching frequencies of occurrence were used as stimuli. To evaluate the effect of the initial segment, the initial consonant of each syllable always formed the first gate, with later gates formed by 40ms increments. Results showed that Tone 1 has a significantly earlier Isolation Point (IP) than Tone 4, which has an earlier IP than Tones 2 and 3. Sonorant-initial syllables have an earlier IP than obstruent-initial syllables, but further analyses of covariance indicated that IP covariates with the duration of the initial consonant. The tone responses proposed by the participants before reaching the IP were cross-examined with the acoustic features of the four tones. The results indicated that high register cues are more prominent than low register cues, as high tones were never misidentified as low tones. Moreover, contour information outweighs low register cues, as low-onset tones were sometimes misidentified as high-onset tones with which they share similar contours. These results provide more detailed temporal information about tone processing for Mandarin.

1. Introduction

As a phonemic feature, tone plays an important role in lexical processing in Mandarin. When listeners hear a word, they need to process both the segmental composition and tone in order to perceive the word correctly. Tones therefore play an important role in isolating the target token from possible segmental homophones. Previous research has shown that the most important acoustic cues for Mandarin tones are F0 height, F0 shape, and F0 differences between the onset and turning point of the tone (especially for Tone 2 and Tone 3). Duration cues such as the overall tone duration as well as the timing of the turning point have also been shown to affect the perception of tones. Given the acoustic differences between the four Mandarin tones, it is necessary to determine how much and what kind of acoustic information is required for listeners to perceive the tones correctly.

1.1. The gating paradigm

In gating experiments, participants are presented with a spoken language stimulus (phone, syllable, word, phrase, or sentence, etc.) in segments of increasing duration, and are then asked to propose the word presented and give a confidence rating (Grosjean, 1996). The increment size is consistent across the stimuli (usually between 20-100ms, or a fixed percentage of an individual word). Three sets of data are usually collected in this type of study. 1) Isolation point (IP) – the size of the segment needed to correctly identify the stimulus without further changes. 2)

Confidence rating – the rating at each segment. 3) Proposed responses – subjects’ responses at each gate before the isolation point. This paradigm allows for precise controls of the acoustic-phonetic information of the stimuli presented to the subjects. As a result, it can examine the moment-to-moment recognition process and evaluate the amount of acoustic-phonetic information required to identify the stimulus.

1.2. Acoustic correlates of Mandarin tones

The F0 contours of the four Mandarin tones produced in isolation are provided in Figure 1. Tones are transcribed as moving within a pitch range from low, numerically denoted as “1”, to high, denoted as “5”. Tone 1 is transcribed as 55 (high-level), Tone 2 as 24 (low-rising), Tone 3 as 213 (low-dipping) and Tone 4 as 51 (high-falling).

Research by Gandour (1983) has shown that five aspects are relevant for tone perception: 1) average F0/F0 height; 2) F0 contour; 3) F0 slope; 4) extreme endpoints; and 5) tone duration. Previous research has shown that the primary acoustic parameters of Mandarin tones are F0 height and contour shape (Howie, 1976). Duration also differs among the four tones: Tone 2 and Tone 3 are the longest, while Tone 4 is the shortest (Nordenhake and Svantesson, 1983). Moore and Jongman (1997) have also shown that Tone 2 has an earlier turning point and a smaller F0 change between the onset and turning point than Tone 3.

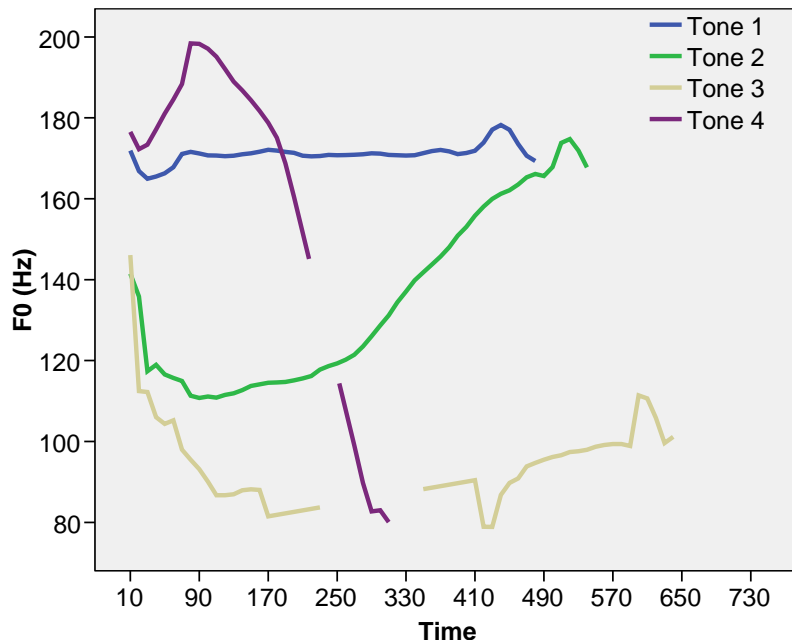


Figure 1. Four Mandarin lexical tones on the syllable *ma* produced by a male native speaker.

1.3. Previous gating studies on Mandarin tones

Lee (2000) used the gating paradigm to explore the lexical competition between different types of syllables by comparing: 1) syllables with tonal minimal pairs to those without; 2) syllables with different numbers of tonal minimal pairs; 3) syllables with similar tones (i.e., Tones 2 and 3) to those with dissimilar tones (i.e., Tones 3 and 4); and 4) syllables with sonorant

onsets to those with obstruent onsets. Three sets of data were collected: the tone isolation point – the point at which the target tone was correctly identified without further changes; the word isolation point – the point at which the target syllable was correctly identified without further changes; and the word recognition point – the point at which the target syllable was correctly identified and the confidence rating reached at least 8 on a 10-point scale. Twenty native speakers participated in the experiments. They were asked to propose (in writing) what word, based on the segment presented to them, they thought they had heard, and give a confidence rating for their answer.

Lee (2000)'s results indicated that tone isolation, word isolation and word recognition points were all earlier for words without tonal minimal pairs. However, no differences were found across stimuli with one, two, and three tonal minimal pairs. In terms of tone similarity, the tone isolation point was consistently earlier for words without tonally similar minimal pairs, but no differences were found in word isolation point or word recognition point. The accuracy rate for tone identification in the initial gate formed by the onset consonant was higher for sonorant onset than obstruent onset.

Wu and Shu (2003) also adopted the gating paradigm in their work on Mandarin tone processing. 120 Mandarin monosyllables were tested on 47 subjects. The gates were constructed with 40ms increments and were presented in a duration-blocked format. The subjects had to write down the character and give a confidence rating for the judgment on a piece of paper. They analyzed the isolation point (IP) of all stimuli, the IP of each tone, and the errors generated by onset, rime and tone.

Their results showed that the IP is the longest for Tone 2; no difference was found between Tones 1, 3, or 4. They also analyzed errors and distinguished errors as coming from either the onset, rime, or tone. It was shown that after the fifth gate (200ms), participants could correctly identify the entire target syllable. They also showed that Tone 1 and Tone 4 were most likely to be mistaken for each other. However, Tone 2 and Tone 3, which have been shown to be similar to each other acoustically (Moore and Jongman, 1997), were not easily mistaken for each other.

There are two methodological problems in Wu and Shu (2003)'s study. First, they did not control for the frequencies of occurrence of the target syllables across different tones in their stimuli. Second, they did not use tone quadruplets to control the segmental composition of the stimuli. Therefore, their result on the processing of different tones might have been confounded with the frequency effects as well as effects from the segmental composition of the stimuli.

1.4. The current study

The present study proposes a revised methodology that provides a better control of frequency of occurrence and segmental composition. Our goal is to investigate the amount of tonal information needed to correctly identify the target tone. This includes the tone duration required from the onset of the token as well as the acoustic cues the listeners adopt during processing. In the meantime, it will also allow us to systematically investigate the effect of sonorancy of the initial consonant on tonal identification.

Figure 2 simulates the hypothesized process of Mandarin tone identification. As shown in the figure, we hypothesize that the four tones will first be distinguished as two groups based on onset tone height. Between the two tones that start with a high register, Tone 1 can be identified earlier than Tone 4, as contour shapes require longer duration to be perceived (Black, 1970;

Greenberg and Zee, 1979). For tones that start with a low register, Tone 2, which has an earlier turning point and a smaller F0 change between the tone onset and the turning point, is predicted to require a shorter duration to identify than Tone 3.

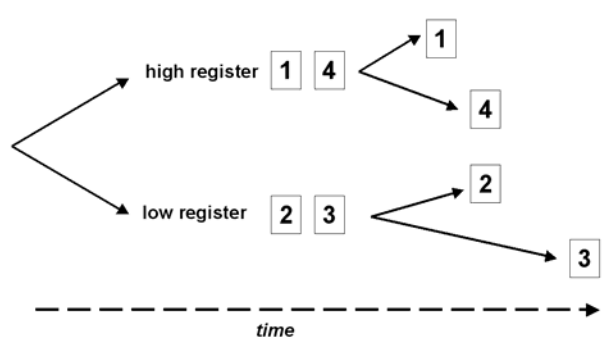


Figure 2. Hypothesized process for Mandarin tone recognition.

In addition, we hypothesize that a sonorant initial can provide acoustic information and subsequently trigger an earlier isolation point as compared to an obstruent initial.

2. Methodology

2.1. Stimuli

The stimuli consist of 8 tone quadruplets, each containing the same segmental composition but different tones. The stimulus list is given in Figure 3. Four quadruplets have a CV structure, while the other four have a CVN structure. The frequencies of occurrence were matched across the four different tones using Da (2007)'s corpus.¹ The design includes two within-subject factors: Tone (1, 2, 3 and 4) and Initial consonant (sonorant and obstruent).

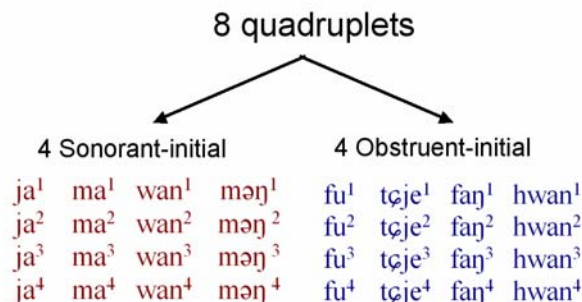


Figure 3. Stimuli- two initial consonant types (sonorant and obstruent).

¹ Construction of the stimuli for Mandarin tone gating is not an easy task. Ideally the frequencies of occurrence for the syllables used in each quadruplet need to be matched. In addition, each syllable in Chinese usually has a number of homophones, each of which has a unique character. For example “有” [y⁴]- the first syllable of the first author’s given name, has 107 homophones. Consequently, syllable frequency, instead of character frequency was used for frequency control. This is because when the stimuli are presented to the subjects auditorily, it is not apparent which character will be activated. This is based on the assumption that when the subjects listen to the stimuli, all homophones will be activated for that pronunciation.

The stimuli were recorded in an anechoic chamber at the University of Kansas by a male native Mandarin speaker. The recording was then transferred to PRAAT for editing in the Phonetics and Psycholinguistics Laboratory at the University of Kansas. The initial consonant of each syllable always formed the first gate. The following gates were formed in 40ms increments starting from the onset of the rime in each syllable. Figure 4 illustrates the complete gating sequence for the Mandarin word “husband” [fu¹].

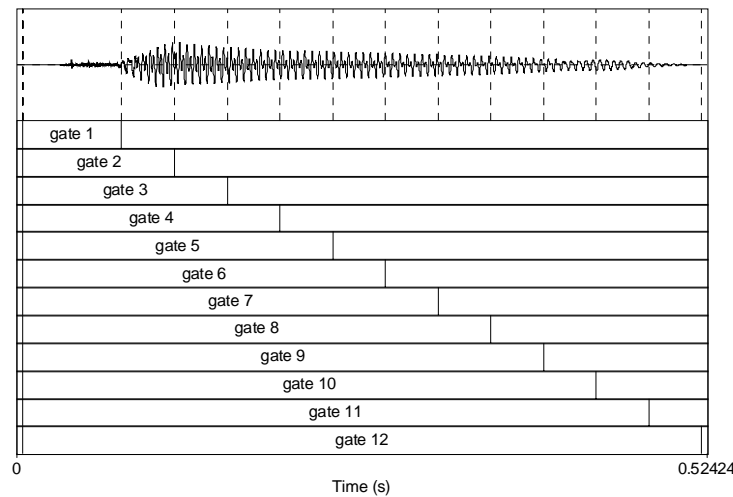


Figure 4. The gating sequence, illustrated by the word [fu¹] “husband”. The first gate includes only the onset consonant. Following gates were constructed in 40ms increments. The last gate always contains the entire syllable.

2.2. Participants

Twenty-eight adult native Mandarin speakers from Beijing were tested at Peking University. They were paid for their participation.

2.3. Experimental procedure

The subjects were tested individually in a quiet room using the SuperLab program (Cedrus). The experiment began with an instruction (recorded by the same speaker who recorded the stimuli), which explained to the subjects that their task was to identify the tone for each gated stimulus and provide a confidence rating on a scale of 1 to 7 for their response by pressing the corresponding buttons on a keyboard. The stimuli were presented in a duration-blocked format, in which subjects heard the first gates of all stimuli, then the second gates, etc.

2.4. Data processing

The following data were collected from the subjects’ responses: 1) isolation point (IP) – the point at which the target tone was correctly identified without further changes; 2) proposed responses before reaching the IP for each tone.

3. Results

3.1. Isolation Point (IP)

The IP results are given in Figure 5. A 4 (Tone) \times 2 (Initial consonant) analysis of variance (ANOVA) of the IP showed a main effect of Tone [$F(3, 864)=114.30, p<.001$]. A posthoc analysis indicated that the IP for Tone 1 was earlier than that for Tone 4, which in turn was earlier than the IPs for Tone 2 and Tone 3. There was no difference between the IPs for Tone 2 and Tone 3. The main effect of Initial consonant is also significant [$F(1, 864)=107.83, p<.001$]; sonorant-initial syllables have an earlier IP than obstruent-initial syllables. The interaction between the two main effects is not significant [$F(3, 864)=2.19, p=.087$].

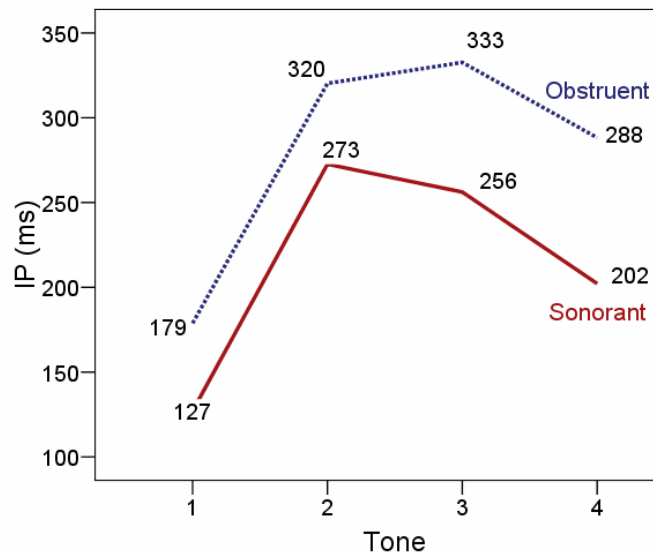


Figure 5. Isolation points across four tones and two types of initial segments.

To further examine whether the earlier IP for sonorant-initial syllables was caused by the sonorancy of the initial segment or a shorter gate 1 duration, we measured the duration of the initial consonant (gate 1). Results from a 4 (Tone) \times 2 (Initial consonant) ANOVA showed that the main effect of Tone was not significant [$F(3, 24)=.44, p=.728$], but there was a main effect of the Initial consonant [$F(1, 24)= 19.35, p< .001$]; sonorant-initial syllables had a shorter gate 1 duration than obstruent-initial syllables. The interaction is not significant [$F(3, 24)=.48, p=.697$]. The gate 1 duration results are given in Figure 6.

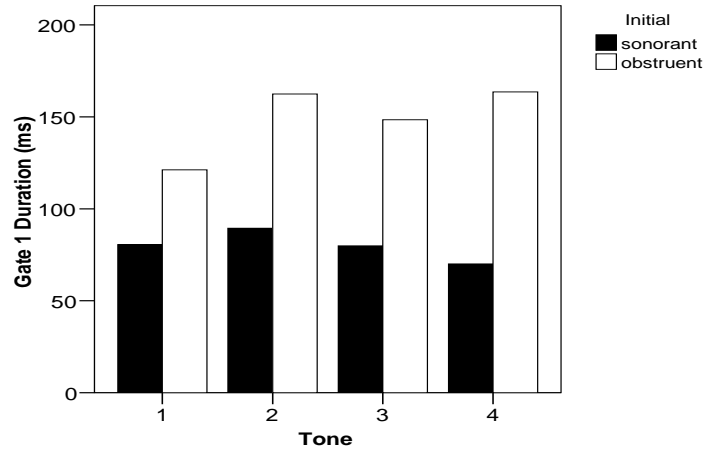


Figure 6. Gate 1 duration across four tones and two types of initial segments.

To evaluate the effect of gate 1 duration on the IP, we further conducted a 4 (Tone) \times 2 (Initial consonant) analysis of covariance (ANCOVA) with gate 1 duration as a covariate. The main effect of Tone was still significant [$F(3, 863)=99.65, p<.001$], and posthoc analyses indicated that Tone 1 had an earlier IP than Tone 4, which had an earlier IP than Tone 2 and Tone 3, and there was no difference between Tone 2 and Tone 3. The main effect of Initial consonant was not significant [$F(1, 863)=.006, p=.939$]. The interaction was also not significant [$F(3,863)=1.92, p=.124$]. This analysis showed that when the factor of gate 1 duration was excluded, there was no significant difference in IP between sonorant-initial and obstruent-initial syllables. The IP results excluding gate 1 duration are given in Figure 7.

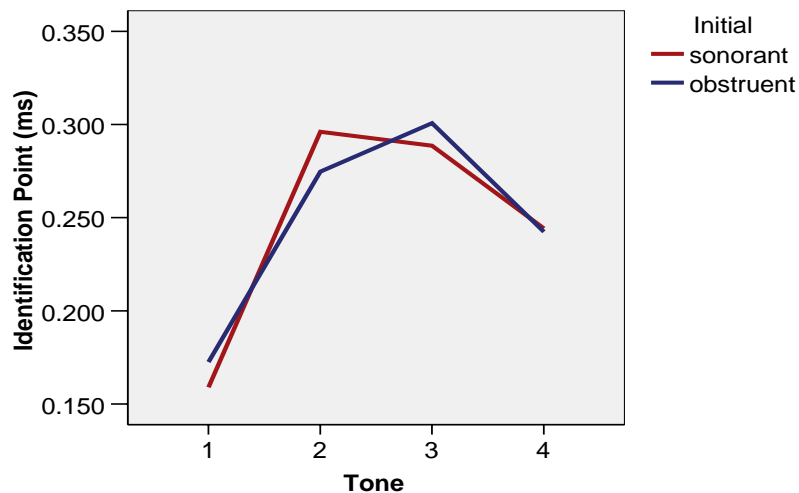


Figure 7. Isolation points across four tones and two types of initial segments (with gate 1 duration excluded).

3.2. Accuracy rate at gate 1

To further investigate the effect of sonorancy on tonal identification, we calculated the accuracy rates for tonal identification at gate 1 for all stimuli. These accuracy rates are given in Figure 8. Results from a 4 (Tone) \times 2 (Initial consonant) ANOVA showed that, unsurprisingly, there was a main effect of tone [$F(3, 888)=44.78, p<.001$], and posthoc analyses indicated that the accuracy rate was the highest for Tone 1, followed by Tone 4 and then Tone 2 and Tone 3. But the main effect of the Initial consonant was also significant [$F(1, 888)=19.53, p<.001$]; the accuracy rate was higher for sonorant-initial syllables than obstruent-initial syllables.

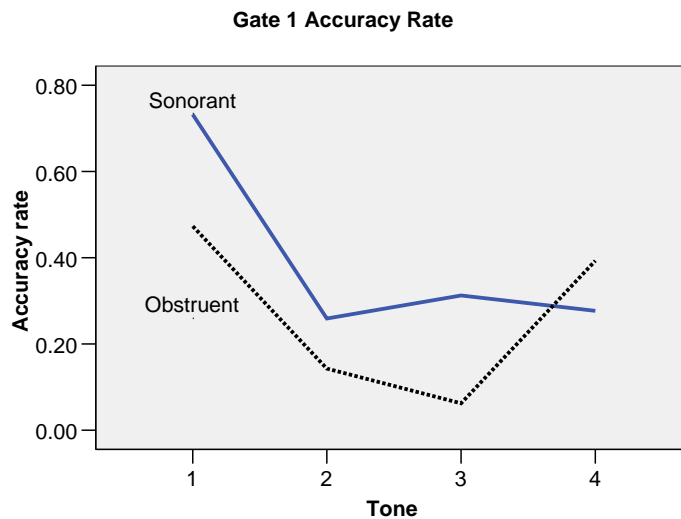
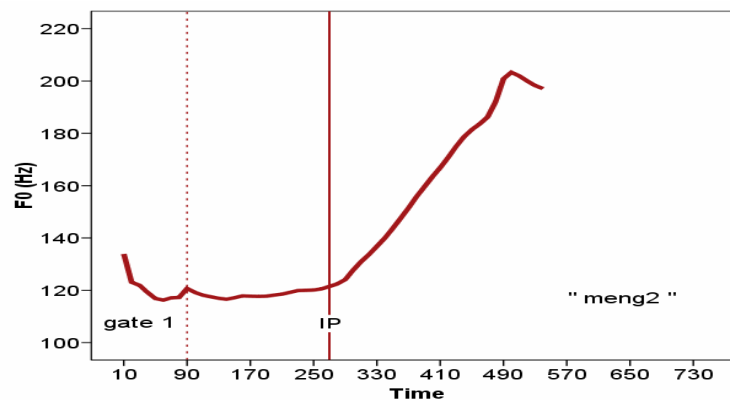


Figure 8. Gate 1 accuracy rate across four tones and two types of initial segments.

To further illustrate this point, sample F0 contours of the sonorant-initial syllable *meng2* and the obstruent-initial syllable *fang2* were plotted in Figure 9. We can see that although the IP for *meng2* is earlier than that for *fang2*, this difference is largely due to the difference in the duration of the initial consonant (gate 1).



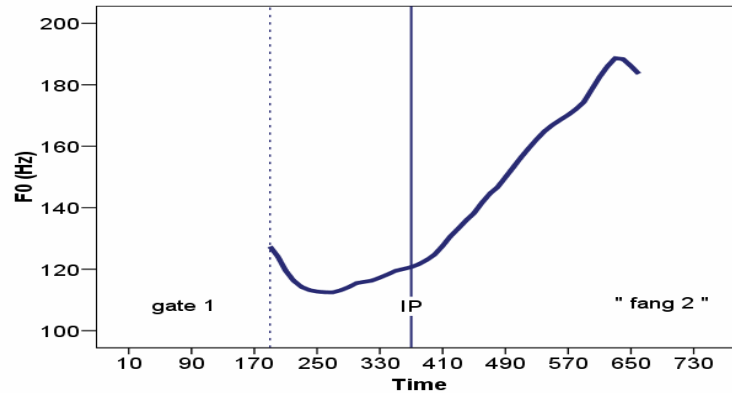


Figure 9. Sample F0 contours for a sonorant-initial syllable *meng2* and an obstruent-initial syllable *fang2*. The dotted line indicates the gate 1 boundary and the solid line shows the IP.

Our ANCOVA results for the IP and the ANOVA results for the accuracy rate indicate that although the sonorancy of the initial consonant does not necessarily trigger an earlier IP, it does contribute to tonal identification in boosting the accuracy rate at gate 1. The fact that it does not trigger an earlier IP indicates that in order for the listener to confidently identify a tone, a certain amount of duration from the vowel is necessary, as it provides clearer acoustic cues for F0.

3.3. Analysis of tonal confusion before IP

We further examined the responses listeners provided before the IP to investigate the possible cues that the listeners used in making their judgments. Figure 10 shows the histograms of tone responses at gates 1-9 for Tone 1 and Tone 4. The y-axis represents the number of times a particular tone was given as the response. For later gates of the stimuli, subjects correctly identified the target tone with a close-to-100% accuracy rate. We thus do not report these later gates in the histograms.

For Tone 1 tokens (Figure 10a), subjects start reaching a high accuracy rate from gate 3. Errors made before gate 3 are mostly misidentifications as Tone 4. This may be due to the similarity between the initial contours of Tone 1 and Tone 4 in both the register and the tone shape. Correspondingly, Tone 4 (Figure 10b), before reaching a high accuracy rate at gate 5, was often misidentified as Tone 1. The reason that it takes longer for Tone 4 to be correctly identified than Tone 1 may be that at earlier gates, the tone duration is not long enough for the subjects to perceive the falling contour, thus causing a level tone perception.

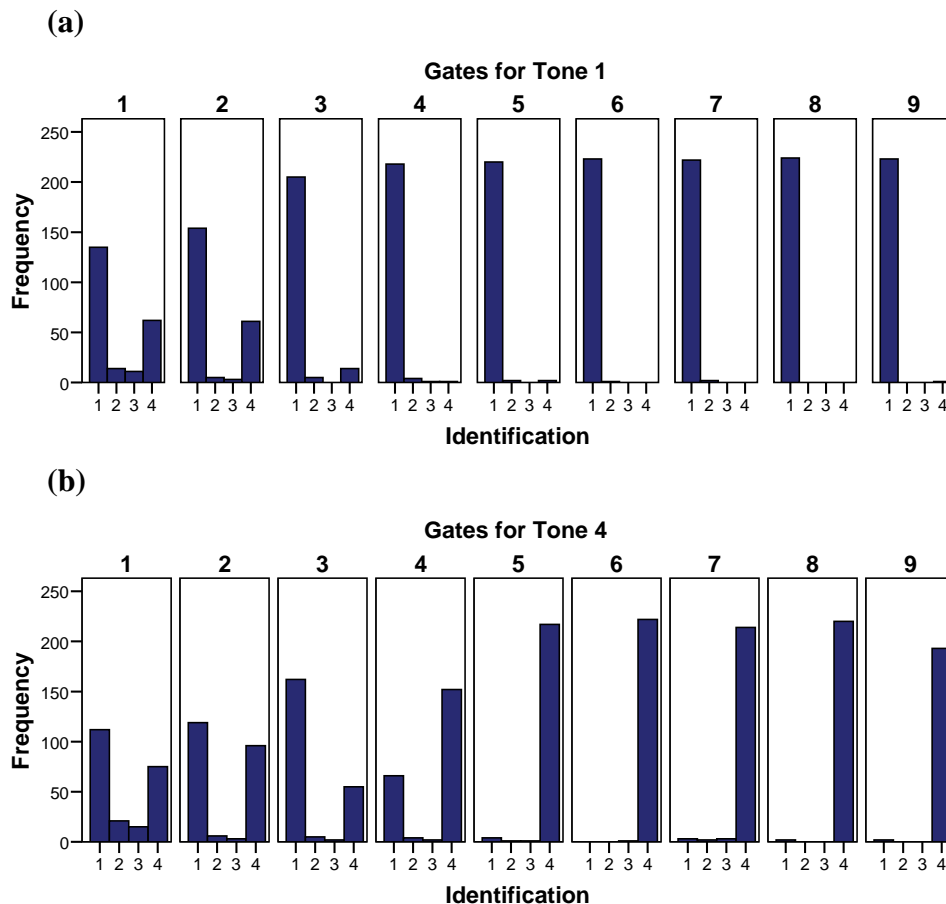


Figure 10. Histograms of tone responses at gates 1-9 for Tone 1 (a) and Tone 4 (b).

The histograms for Tone 2 and Tone 3 responses for gates 1-9 are given in Figure 11. Similar to the results found in Wu and Shu (2003), there was a high percentage of Tone 1 responses at earlier gates for Tone 2 and Tone 3. We propose that this is again due to the short duration of the presented segment, which does not warrant a contour tone perception. Tone 1, being the only level tone in the language, then becomes the most common response at earlier gates for all four target tones. Interestingly, instead of being confused with Tone 3, Tone 2 tokens at earlier gates received significant Tone 1 responses. Closer examination of Tone 2 and Tone 1 showed that at early gates, these two tones, although different in register, share very similar tonal contours (cf. Fig. 1).

The first gate of Tone 3 was also often misidentified as Tone 1, presumably due to the short duration. Starting from gate 2, subjects had significantly more Tone 3 responses. Interestingly, in gates 4-7, Tone 3 tokens were sometimes misidentified as Tone 4. We surmise that since Tone 3 is the only tone in the low register region, listeners may have taken advantage of the low register and identified them as Tone 3 at the very beginning; but when enough duration was heard, which warranted a falling tone perception, a Tone 4 perception was triggered.

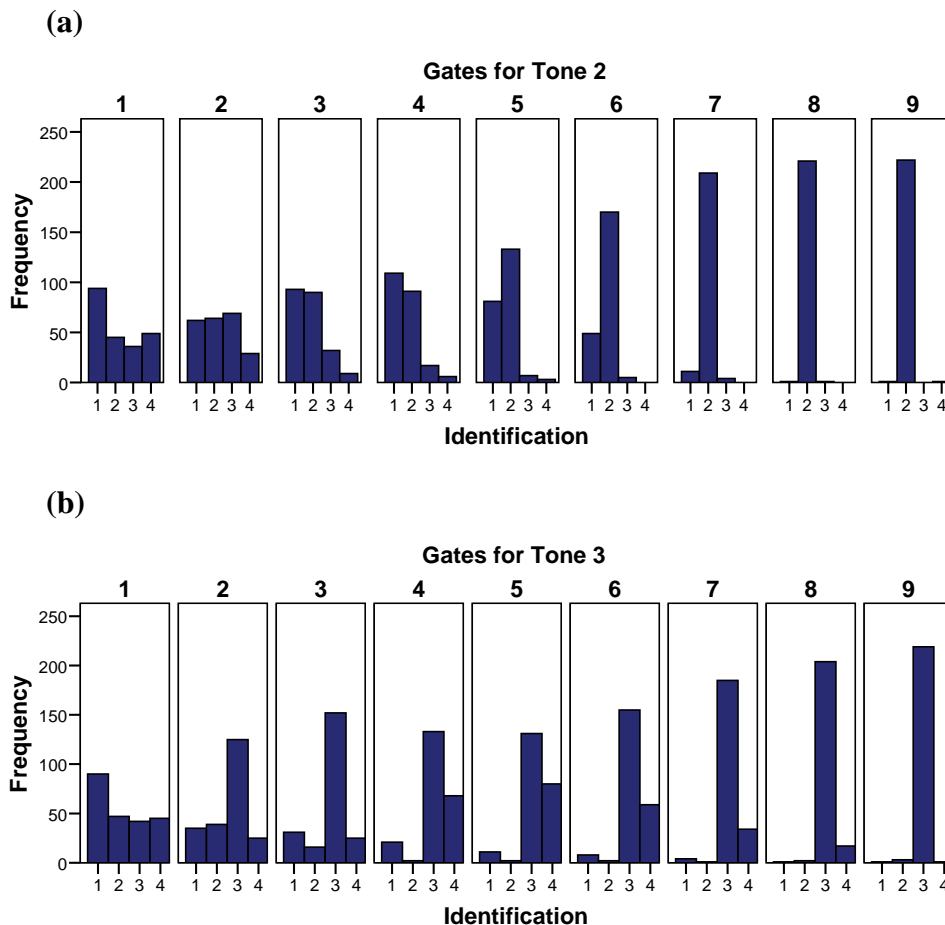


Figure 11. Histograms of tone responses at gates 1-9 for Tone 2 (a) and Tone 3 (b).

The confusion analysis suggests that among the acoustic cues for Mandarin tones, listeners are the most sensitive to a high registered pitch at the beginning of a tonal stimulus, as evidenced by the earlier IPs for Tone 1 and Tone 4, which start with a high pitch. The contour perception requires a significant duration of the vowel that carries the contour, as shown by the common misidentification of Tone 4 as Tone 1 at early gates. An acoustic low pitch seems to carry a less significant perceptual weight than an acoustic high pitch at tonal onset, as evidenced by the common misidentification of Tone 2 (low pitch onset) as Tone 1 (high pitch onset) at early gates; the low pitch also carries a less significant perceptual weight than a contour pitch at tonal onset, as the speakers sometimes identified the low falling pitch at the beginning of Tone 3 as the high falling pitch of Tone 4.

Our tonal confusion data are inconsistent with the general understanding that among Mandarin tones, Tone 2 is more likely confused with Tone 3, and Tone 1 is more likely confused with Tone 4. This may be due to the fact that the cues listeners focus on during the initial unfolding of the tone are different from the ones they use once the entire tone has been presented.

4. Discussion and conclusion

The current study establishes that a timing difference exists in the processing of different Mandarin tones. The isolation point is the earliest for Tone 1, followed by Tone 4, which is then followed by Tone 2 and Tone 3. The IP for sonorant-initial syllables is earlier than that for obstruent-initial syllables, but this difference is likely due to the shorter duration of initial sonorants than initial obstruents, not their difference in sonorancy per se. Despite its lack of temporal effects, the sonorancy of the initial consonant does contribute to the identification of tone, in that it boosts the accuracy rate of identification at gate 1, which is solely composed of the initial consonant.

Based on the confusion analysis before the isolation point, a hierarchy of cues at the onset of tonal identification was also found: high > contour > low. High-onset tones, regardless of contours, were not misidentified as low-onset tones; but low-onset tones were sometimes misidentified as high-onset tones due to their contour shapes.

In sum, our study provides more detailed temporal information about tone processing for Mandarin. With a better understanding of the timing of identification and acoustic cues that listeners rely on at the onset of identification, it will assist in the further refinement of temporal precision in future processing studies of Mandarin tones.

References

- Black, J. W (1970). The magnitude of pitch inflection. *Proceedings of the 6th International Congress of Phonetic Sciences*. Prague: 177-181.
- Da, J. (2007). *Chinese text computing corpus*. <http://lingua.mtsu.edu/chinese-computing/copyright.html>
- Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11: 149-175.
- Greenberg, S. and E. Zee (1979). On the perception of contour tones. *UCLA Working Papers in Phonetics* 45: 150-164.
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, 28 (4): 267-283.
- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.
- Lee, C-Y. (2000). *Lexical tone in spoken word recognition: a view from Mandarin Chinese*. Doctorate Dissertation, Brown University.
- Moore, C. B., and Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America*, 102: 1864-1877.
- Nordenhake, M. and J. O. Svantesson. (1983). Duration of standard Chinese word tones in different sentence environments. *Lund University Working papers in Linguistics* 25: 105-111.
- Wu, N. and Shu, H. (2003). The gating paradigm and spoken word recognition of Chinese. *Acta Psychologica*, 35 (5): 582-590.

Author contact information:

Yuwen Lai: yuwen@ku.edu

Jie Zhang: zhang@ku.edu