Unmerging T2 and T3 in Urumqi accented Standard Mandarin

Xinlu Yang\textsuperscript{1,2}, Jie Liang\textsuperscript{1}, Marjoleine Sloos\textsuperscript{3}

\textsuperscript{1}School of Foreign Languages, Tongji University, China
\textsuperscript{2}College of Foreign Languages, Xinjiang Normal University, China
\textsuperscript{3}Aarhus University, Denmark

xlyangxjnu@sina.com, liangjie56@163.net, marj.sloos@gmail.com

Abstract

The Urumqi Mandarin dialect has three tones, whereas Standard (Beijing) Mandarin has four tones. Recently, younger speakers in Urumqi began to use Standard Mandarin in daily conversation. We investigate how they acquire the tones in Standard Mandarin, especially the rising tone (T2) which lacks a counterpart in Urumqi Mandarin. We will show that these younger speakers initially use Urumqi tones in Standard Mandarin (but for different lexical sets). At that stage, the production of T1, T3 and T4 approximates Standard Mandarin relatively well. However, T2 is merged with T3. In the second stage, unmerger of T2 and T3 occurs, which is characterized by a distinction between early (T2) and late (T3) low turning points in the dipping tone. In the third stage, further disambiguation occurs by lowering the lowest turning point in T3. However, the final rise of T2 is still missing. We assume that raising the final part of T2 will be the final step to be taken in the near future.

Index Terms: tone, intonation, Mandarin, Urumqi

1. Introduction

Urumqi is a Mandarin dialect of the Lanyin branch, spoken in northern Xinjiang. In recent years, younger speakers of Urumqi dialect started to use Standard (Beijing) Mandarin in their daily life [1]. Impressionistically, their production sounds clearly accented, which seems to be caused by the prosody. In this contribution, we investigate the acquisition of the Standard Mandarin lexical tones produced among younger speakers.

Comparing Urumqi tones with Standard Mandarin tones (using Chao’s five-level numerical scale), Urumqi has three lexical tones: 44, 52 or 51, and 312 or 213 [2,3]. Standard Mandarin has four tones, usually transcribed as 55, 35, 214, and 51 (see also Table 1).

Table 1. Tonal systems of Urumqi and Standard Mandarin.

<table>
<thead>
<tr>
<th>Tone</th>
<th>Urumqi</th>
<th>Standard Mandarin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>44</td>
<td>55</td>
</tr>
<tr>
<td>2</td>
<td>51/52</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>312/213</td>
<td>214</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>51</td>
</tr>
</tbody>
</table>

As Table 1 shows, the falling tones in Urumqi and Standard Mandarin are comparable, but they occur in different lexical sets. The dipping tone in Urumqi is comparable to the dipping tone in Mandarin, but is also applied to different sets. Crucially, in Urumqi, the lexical sets that have T2 and T3 in Mandarin are merged: both are produced with the same falling tone 51/52 [2,3].

How do the younger speakers acquire the lexical tones of Standard Mandarin? Is it the case that the complete tonal system is acquired at once? This seems unlikely, given the `accent’ of younger Urumqi speakers. Is it the case that the younger speakers gradually change the tones? Or do they shift the dipping and falling tones at once—which would immediately raise another question: how do they acquire T2 (the rising tone)?

To investigate the acquisition of T2 in younger Urumqi speakers of Standard Mandarin, we conducted an experiment in which we compared the tonal systems of the older and younger speakers in Urumqi. Since T2 is the rising tone in Standard Mandarin, it is likely that this contour is acquired earlier in question intonation than in statement or citation intonation. We therefore considered, besides the citation tones, the tones in statement and question intonation. We will show that the older speakers retained the three Urumqi tones, but the younger speakers use an interesting mixture of Urumqi and Mandarin. They shifted the falling T2/3 and dipping T4 Urumqi tones to Standard Mandarin, resulting in falling T4 and dipping T2/T3. Thus, T2 and T3 received another tone, but still they were merged. In the following stage, gradual unmerger of T2 and T3 occurred. This unmerger involved not only contour change (as should be expected) but also a timing difference of the low turning point.

Summarizing, in this contribution, we investigate the following research questions:

i) what is the tonal system of older and younger Urumqi speakers?
ii) if tone shift occurs, is it gradual or abrupt?
iii) specifically, how do younger Urumqi speakers acquire T2?
iv) is there a difference in realization of T2 in citation form, statement, and question intonation?

2. Methodology

2.1. Stimuli

The stimuli were provided through a reading list consisting of two parts. The first part (citation tones) contained 16 Mandarin monosyllabic words, taken from The Dictionary of Urumqi Dialect [3]. These words were all combined with the three lexical tones in Urumqi and the four lexical tones in Mandarin, as illustrated in Table 2.
Table 2. Stimuli for tone production task.

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>呀 [tei]</td>
<td>魏 [tei\’an]</td>
<td>票 [pi3]</td>
<td>票 [p^i3]</td>
</tr>
</tbody>
</table>

The second part of the reading list included 16 sets of utterances, derived from the 16 monosyllabic words in Table 2. Each set contained two types of utterances: one statement and one syntactically unmarked yes-no question. The target words were embedded in 16 mini-dialogues as in (1).

(1) Context:

那是个啥东西？

na shige sha dongxi?

“What’s that?”

Target word with question intonation:

鸡?

ji?

“Chicken?”

Target word with statement intonation:

鸡。

ji

“Chicken.”

2.2. Subjects

Sixteen speakers, born and raised in Urumqi, participated in the study. Eight older speakers were third-generation Urumqi citizens and eight younger speakers were second-generation Urumqi citizens. The age of the older speakers ranged from 50 to 60 years and the age of the younger speakers ranged from 19 to 21 years. They were paid for their participation. None of the participants reported any hearing, vision, or reading deficiencies.

2.3. Procedure

During individual recording sessions, the subjects were seated at a desk in a quiet room at Xinjiang Normal University. The participants produced the monosyllabic stimuli by reading aloud slides presented by Microsoft PowerPoint software. The speech was recorded with the microphone of a Sennheiser PC 300 high-quality headset. If the speaker failed to read a word aloud, he or she was prompted immediately by the experimenter (first author) to repeat the attempt.

2.4. Acoustic analyses

The onset consonant of each syllable is an obstructed and doesn’t carry any distinctive information with respect to lexical tone, so we analysed only the rhyme. Each monosyllabic tone production was manually segmented into onset and rhyme by using a Praat script [4]. The script measured the F0 at 10 equidistant points of the rhyme. To exclude transitions, we only used points 2-9 for the analyses.

Subsequently, in order to normalize for gender differences, the F0 was converted into a semitone scale (a psycho-acoustic scale equal to perceptual intervals [5]). To visualize the differences between the groups of speakers (based on gender and age), we computed the pitch range for each speaker (the highest minus the lowest pitch) and subtracted that from the average pitch range of all speakers.

3. Results

3.1. Pitch contours of the citation tones

To see if the younger speakers used four tones and the older speakers used three tones, we compared the average pitch contours for each tone for the older and the younger speakers. Figure 1 shows that the tones of the older speakers are 44, 51, and 213 (cf. the literature cited above). Figure 1 shows that the younger speakers produced tones similar to Standard Mandarin. Moreover, the four tones produced by the younger female speakers are more differentiated than the four tones produced by the younger male speakers. T2 and T3 as produced by the younger males are not very different from each other.

![Figure 1](image-url)

Figure 1: Pitch contours of the citation tones for older and younger speakers.

Table 2 (at the end of this subsection) provides the tones in Chao numbers.

Since individual differences in tone production often occur, we examined the individual averages of the tones for the younger speakers. T1, T2 and T4 turned out to be quite similar across the speakers, but T3 was clearly variable. Consequently, the differentiation between T2 and T3 differs per speaker. To visualize this differentiation between T2 and T3, we present the individual average contours of these two tones by the younger speakers. The results are shown in Figures 2-4.

Figure 2 shows the contours of T2 and T3 in the citation forms. No differentiation between T2 and T3 among the male speakers can be found: both are slightly dipping. However, the female speakers disambiguate T2 and T3 by lowering the lowest turning point in T3. Figure 3 provides the contours of
T2 and T3 in statement condition. Again, no difference between the male speakers can be detected. But here, we find even more disambiguation among the female speakers by a lower turning point in the dipping T3. Figure 4 illustrates the contours of T2 and T3 in question intonation. Male speakers seem to disambiguate these tones more than in citation and statement condition. Female speakers approach Standard Mandarin pronunciation by the rising contour, but T3—due to the rising question intonation—again turns out to merge with T2 (Figure 4).

Figure 2: Pitch contours of the citation forms of T2 and T3 as produced by the individual younger speakers, divided by age and gender. Each line represents the average contour of a single speaker.

Figure 3: Pitch contours of T2 and T3 in statement condition as produced by the individual younger speakers, divided by age and gender. Each line represents the average contour of a single speaker.

Figure 4: Pitch contours of T2 and T3 in question condition as produced by the individual younger speakers, divided by age and gender. Each line represents the average contour of a single speaker.

From Figure 1, we derive the tonal contours for the older speakers and from the Figures 2-4, we derive the tonal contours for the younger speakers.

Tone 1
Since the falling tone started higher than the high level tone, we analyse T1 as 44, for both the older and the younger speakers.

Tone 2
The older speakers merged T2 and T3, which was produced as 51 by males and 52 by females. T2 was produced as a slight rising tone 23 by younger males and a slight dipping tone 323 by younger females.

Tone 3
T3, as produced by the younger speakers, was marginally different from T2. The male speakers produced T3 as 212 and two females produced T3 as 313.

Tone 4
The older speakers produced T4 as a dipping tone 213. The younger speakers produced T4 as a falling tone, but rather than using the whole pitch range of 51 (like older males and like generally in Standard Mandarin), they produced a high falling tone 53.

The contours for the older and younger speakers as well as Standard Mandarin are summarized in Table 2.

### Table 2. Tonal systems of Urumqi older and younger speakers and Mandarin ((m) refers to male speakers and (f) refers to female speakers).

<table>
<thead>
<tr>
<th>Tone</th>
<th>Urumqi Older</th>
<th>Urumqi Younger</th>
<th>Standard Mandarin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>44</td>
<td>44</td>
<td>55</td>
</tr>
<tr>
<td>2</td>
<td>51(m)/52(f)</td>
<td>23(m)/323(f)</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>213</td>
<td>212/313</td>
<td>214</td>
</tr>
<tr>
<td>4</td>
<td>213</td>
<td>53</td>
<td>51</td>
</tr>
</tbody>
</table>

3.2. Timing of the low turning point

Careful comparison of the contours in Figures 2-4, especially the graphs of the female productions, shows that the lowest
dipping point is aligned early in T2 and late in T3. In order to test whether the low turning point significantly differs for T2 and T3, we conducted a Wilcoxon test for all realizations of T2 and T3 in citation form with the alignment of the low turning point expressed through its location on the 10-point scale as the dependent variable. The differences were significant within the females (W(15) = 182, Z = −3.09, p = 0.001). The female speakers differentiated similarly in statement condition (W(15) = 163, Z = −3.80, p < 0.001) and nearly significantly in question condition (W(15) = 212, Z = −1.96, p = 0.051). No differentiation among the younger male speakers was found. So T2 and T3 are differentiated by the younger female speakers by early and late alignment of the low turning point, respectively, which we will denote as 3233 (T2) and 3323 (T3).

4. Discussion

We compared the production of tones of older and younger speakers in Urumqi and raised the question how younger speakers acquire the rising tone while they accommodate toward Standard Mandarin. Older speakers use Urumqi Mandarin with the following tones: T1 44, T2/3 51 or 52, T4 213. Younger speakers use Standard Mandarin and shifted the falling tone and dipping tone, accordingly. However, the falling tone only uses the higher register and lacks the lowest part. Since Urumqi Mandarin has no rising tone, the younger speakers have to use other strategies to acquire T2 in their Standard variety. Younger females turned out to disambiguate T2 and T3 consistently by early alignment of the lowest turning point in T2 and late alignment of the lowest turning point in T3. In addition, some female participants also showed lowering of the low turning point in T3, approaching Standard pronunciation. Younger males turn out not to distinguish T2 and T3 in production according to our measurements. A perception experiment could give more insight into the distinction in the younger males speech. We leave that to future research.

The variation among the younger speakers in their production of T3 suggests ongoing change, which unmerges T2 and T3. We can distinguish different stages in this change:

i. initially, T2 and T3 probably both had the contour 213—the same as T4 in Urumqi Mandarin.

ii. as can be observed in the younger males’ productions, T2 and T3 were slightly disambiguated into 23 and 212, respectively.

iii. as can be observed in the younger females’ productions, T2 and T3 are further differentiated through early and late alignment of the low turning point, respectively, resulting in 3233 (T2) and 3323 (T3).

iv. as can be observed in the production of two younger females, the low turning point of T3 is further lowered.

Some change that might take place in the future if Urumqi speakers will accommodate further toward Standard Mandarin may involve raising of T2 to the higher register and a steeper rise; stronger rising of the final part of T3; and a steeper fall of T4. T1 is 44 in our speakers, whereas the realization 55 is common in Standard Mandarin. However, 44 is not uncommon, so a change towards 55 is not expected in the short term. The change is summarized in Figure 5.

Figure 5: Tonal change from Urumqi Mandarin to Standard Mandarin. Changes are displayed in bold.

The absence of the high rising contour of T2, the relatively levelled production of T3, and the absence of the lower falling part of T4 in the younger speakers’ production is, we believe, responsible for the perception of strong accent in their variety.

5. Conclusion

Urumqi speakers are currently shifting towards Standard Mandarin. Since the tonal systems in the two varieties are quite different, this change is not straightforward. We distinguish the following stages in this change: tone shift of the falling and the dipping tone, unmerging T2 and T3 by (i) contour differentiation, (ii) differentiation of the alignment of the low turning point in the dipping tone T3 , (iii) differentiation by further lowering of the low turning point in T3. We argued that the change is not yet complete and we expect some modifications to occur in the future.

6. Acknowledgements

Funding was provided by the Ministry of Education in China Project of Humanities and Social Sciences 13YJC740124. We thank Vincent van Heuven for valuable suggestions and Jeroen van de Weijer for comments on an earlier version of this paper.

7. References