How many types of neutral tones in Mandarin: a case study of perception

Bei Yang

Department of East Asian Languages and Literature, University of Wisconsin -- Madison, USA
byang25@wisc.edu

Abstract

This paper explores the environments of tonal neutralization in Mandarin using two perception experiments: Experiment 1 examines whether changing the pitch of the neutral tone can affect how it is perceived; Experiment 2 investigates whether a neutral tone with a lengthened duration can be perceived as a full or neutral tone naturally. Twenty native speakers participated in this study. The results indicate that pitch plays an important role during perception. Therefore, four full tones become indistinguishable in four phonetic environments and there are four categories of neutral tones. The results are similar to the tone sandhi in Wu dialects, giving rise to some interesting comparisons.

Index Terms: neutral tone, tone sandhi, perception

1. Introduction

Phonological neutralization is well documented across languages, such as a word-final devoicing in German and Russian [3, 13, 14]. The term describes "forms which are distinguishable phonetically and phonologically in certain contexts and/or levels of representation are under certain other well-defined circumstances totally indistinguishable at the level of phonetics" ([6], 1985: 265). The present paper focuses on the neutral tone (T0) in Mandarin.

While the shì in chāshì 左使 (position) and the shì in chāshì 差事 (errand) differ in their underlying representation in terms of tones, the surface forms are identical in both words. Actually, there are four full tones in Mandarin. In some di- or tri-syllabic words, the tone carried by the second syllable is neutralized when the latter conveys less semantic information than the former does. However, the neutral tone is not always related to a weak meaning. It could also occur in a disyllabic word in which the second syllable has the same semantic load as the first one.

In German, word-final devoicing is indistinguishable in a two-way contrast during neutralization, i.e. voiced vs. voiceless. However, four full tones are indistinguishable when the tonal neutralization occurs in some contexts in Mandarin. As a result, the neutralization of the tones involves a four-way contrast and the neutral tone is different from the original full tone that is also called the base tone. This suggests an important question: in which phonetic environments are tones neutralized? In the literature, many features of the neutral tone have been studied including intensity, duration, pitch, as well as finals that carry the neutral tones and its relevant initials. The studies mentioned below are based on native speakers’ production.

The neutral tone is called qinsheng (light sound) in Chinese because it is perceived to be softer than full tones. Chao [1], Lin and Yan [10] all posited that the neutral tone has low intensity. However, further studies indicated that the intensity of a neutral tone is not necessarily less than that of a full tone [11]. Lee and Zee [8] found that intensity is covariate of pitch, demonstrating that intensity is not an important feature of the neutral tone. However, studies on duration and pitch of the neutral tone revealed that duration and pitch are the salient features.

Early acoustic studies showed that the salient feature of the neutral tone is its shorter duration. Acoustic analysis also revealed that the duration of the neutral tone is much shorter than that of full tones. Some studies [1, 5, 9] found that the duration of a neutral tone is around 50%-60% of the full tones. It is reasonable that tones of shorter duration require less energy, making listeners perceive the neutral tones as less stressed than full tones.

Moreover, some researchers found that the neutral tone carried different pitch values. It could be a falling or rising contour [3, 5, 8] or a level contour [4]. If the tone preceding a neutral tone is T1, T2 or T4, the neutral tone can be perceived as a falling tone at mid, high or low register respectively. If the preceding tone is T3, the neutral tone is perceived as a rise first, and then a fall. As a rule, whatever the contour of the base form of the neutral tone is, the neutral tone’s contour is determined by the preceding full tone. In fact, Chen & Xu [5] observed three consequent neutral tones in context and claimed that the neutral tone has a falling tendency.

Meanwhile, it is also found that the neutral tone influences the initial and the final in the same syllable [2, 17]. For example, voiced consonants could become voiceless consonants; the main vowel in the final could become a schwa sound; the nasal or the whole final could be lost, such as dōuf (bean curd) and wōm (we). These could also explain why the syllable carrying the neutral tone is perceived as weaker than the syllable carrying the full tone. However, initials and finals are not always influential. Therefore, they could not be regarded as general features of the neutral tone.

In conclusion, shorter duration and dependent pitch are two salient features of the neutral tone. The former is different from the full tones, while the latter is partially similar to the full tones. Thus, the problem is that the phonetic environments of the tonal neutralization in terms of pitch and those in terms of the duration contradict one another.

There are four types of neutral tones according to the phonetic environments in terms of the pitch. The neutral tones after different full tones have different pitch values. Therefore, in terms of the pitch, four full tones are indistinguishable after tone 1, tone 2, tone 3 and tone 4 respectively. The tonal neutralization occurs in four phonetic environments. In terms of the duration, tonal neutralization occurs only in one environment, i.e., four full tones are indistinguishable after any full tones, because of the shorter duration. This involves the nature of the neutral tone.

From the perspective of perception, how many types of the neutral tones can we perceive? In which phonetic environments are the full tones indistinguishable? In other words, should neutral tones be categorized as four weak tones or one weak tone? In order to answer these questions, the perception experiments are conducted in this study to explore the environments of tonal neutralization in Mandarin.
There are two experiments in this study. Experiment 1 explores whether changing the pitch of a neutral tone can affect its perception. Experiment 2 investigates whether a neutral tone with a lengthened duration can be perceived as a full tone or a neutral tone naturally. The purpose of these two experiments is to investigate whether the pitch or the duration plays an important role in tonal perception. If the duration is dominant, tonal neutralization occurs only in one environment. If the pitch is dominant, tonal neutralization occurs in four environments. Otherwise, both the duration and the pitch are important and the neutralization environments will be a little more complicated than the previous two situations.

2. **Experiment 1: Perceptual Experiment on Pitch of T0**

This experiment examines whether pitch is dominant in perceiving the neutral tone.

2.1. **Method**

2.1.1. **Subjects**

Twenty subjects from Mainland China whose native language is Mandarin Chinese participated in this experiment. Some of them speak different dialects, but Mandarin is their primary Chinese language. Their ages range from 20 to 35.

2.1.2. **Stimuli**

Most studies indicate that the pitch contour of the neutral tone depends on the tonal category of the previous syllable (e.g. [1]). Accordingly, four words in which the first syllables carry four different tones are selected (see table 1). 0 stands for the neutral tone, subscript number 1, 2, 3, 4 represent the neutral tone originally come after T1, T2, T3 and T4 respectively.

<table>
<thead>
<tr>
<th>word</th>
<th>Meaning (Base Form)</th>
<th>Pinyin (Sandhi Form)</th>
</tr>
</thead>
<tbody>
<tr>
<td>哥哥</td>
<td>brother</td>
<td>ge1ge01</td>
</tr>
<tr>
<td>葛格</td>
<td>princess</td>
<td>ge2ge02</td>
</tr>
<tr>
<td>葛葛</td>
<td>Gege (name)</td>
<td>ge3ge03</td>
</tr>
<tr>
<td>个个</td>
<td>everyone</td>
<td>ge4ge04</td>
</tr>
</tbody>
</table>

The second syllable ge0 in each word is cut and then combined with the first syllables of the four target words, that is, ge1, ge2, ge3 and ge4. This forms four groups of words (see table 2). Stimuli are synthesized by Praat.

2.1.3. **Procedure**

Subjects were required to listen to the above four groups of words in a carrier sentence: "________ dou1 qu4 shang4ban1 le" (Even _______ went to work).

First, the subjects read a sentence in Chinese, and then they listened to four sentences that had the words with modified tones embedded. After that, they need to select the sentence which they think matches the sentence that they read. For example, the subjects read the sentence “the brother went to work,” and then they listened to following four sentences:

A) ge1ge01, dou1 qu4 shang4ban1 le  
B) ge1ge01, dou1 qu4 shang4ban1 le  
C) ge1ge01, dou1 qu4 shang4ban1 le  
D) ge1ge01, dou1 qu4 shang4ban1 le  

Then, they were asked to select a sentence which they thought matched the sentence they read best.

2.2. **Results**

Most participants selected the original neutral tones matching what they heard. Table 3 indicates that over 85% of participants selected the original neutral tones matching the neutral tones after T1 and T2, and 70% participants selected the original neutral tones matching the neutral tones after T3. As for T4, the results varied. However, nobody selected the third choice, i.e. T0 originally after T3. The reason is that the pitch contours of the neutral tones after T1, T2 and T4 are falls, while that after T3 is a rise (Figure 1). However, we found that most participants still selected the original T0 (D) for the pattern “T4+T0”. Additionally, 10% of participants also selected D for “T1+T0”, since T0 after T1 and T4 are non-high falling tones. 15% of participants also selected A for “T2+T0”, since T0 after T2 and T1 are non-low falling tones.

Table 3. Perceptual results of T0 with modified pitch.

<table>
<thead>
<tr>
<th>Target tones</th>
<th>T1+T0</th>
<th>T2+T0</th>
<th>T3+T0</th>
<th>T4+T0</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (T0 originally after T1)</td>
<td>90%</td>
<td>15%</td>
<td>5%</td>
<td>30%</td>
</tr>
<tr>
<td>B (T0 originally after T2)</td>
<td>0</td>
<td>85%</td>
<td>0</td>
<td>30%</td>
</tr>
<tr>
<td>C (T0 originally after T3)</td>
<td>0</td>
<td>0</td>
<td>70%</td>
<td>0</td>
</tr>
<tr>
<td>D (T0 originally after T4)</td>
<td>10%</td>
<td>0</td>
<td>25%</td>
<td>40%</td>
</tr>
</tbody>
</table>

In Table 3, 30% participants selected the falling T0 (originally after T1 or T4) for “T3+T0” totally. It is possible that some native speakers regard the neutral tone as a falling tone, because there is a tendency that the neutral tone changes to the falling tone [16]. On the other hand, T0 after T1, T3 and T4 are non-high tones.
Although native speakers selected some variations when listening to T0 in different contexts, i.e. after T1, T2, T3 and T4, almost all of them selected the original T0 for "T1+T0", "T2+T0", and "T3+T0", and more native speakers chose the original T0 for "T4+T0". This indicates that the pitch of the neutral tone does affect listeners' comprehension. Tonal pitch plays an important role in the perception of the neutral tone.

3. Experiment 2: Perceptual Experiment on the Duration of T0

The salient feature of the neutral tone is its shorter duration and the pitch variations. Experiment 1 focuses on the pitch, while experiment 2 explores whether the neutral tone whose duration is lengthened can be perceived as a full tone, such as T3, or T4, or a neutral tone naturally.

3.1. Method

3.1.1. Subjects

The 20 subjects participating in this experiment are the same as those in Experiment 1. Only 19 of them completed the test in this experiment.

3.1.2. Stimuli

There are six modified tones and four original tones in carrier sentences in this experiment. In total there are 10 sentences for native speakers to listen to. Sentence 1 to 4 are related to the original neutral tones. Sentences 5-8 are related to the double-duration neutral tones in place of the original neutral tones. Sentences 9-10 have the double-duration neutral tones in place of the full tones, i.e. T3 and T4.

The syllables used here are the same as those in Experiment 1 (see table 1 above). The duration of the tones were synthesized in the following manner. First, the neutral tones which come after T1, T2, T3 and T4 are cut from the sentences. They are labeled as T01, T02, T03, and T04. The four neutral tones are lengthened to double duration via the program Praat. The minimum pitch is set to 50Hz, the maximum pitch is set to 350Hz, and the factor is set to 2. Four tones with double duration of the original neutral tones are synthesized, and are labeled as T01+, T02+, T03+, and T04+.

Then T01+ is inserted into the carrier sentence "ge3ge0 dou1 qu4 shang4ban1 le" (Even Gege went to work) in place of T01. Meanwhile, T02+, T03+ and T04+ are inserted into the other three carrier sentences respectively, instead of T02, T03, and T04. This produced the four sentences with four lengthened neutral tones.

As for the stimuli relevant to full tones, T001 and T002 were selected to substitute the full tones, i.e. T3 and T4 in the stimuli sentences. Some studies (e.g. [1]) found that neutral tones have different pitch contours depending on the tones of the previous syllable. Fig 1 shows four pitch contours of the neutral tone when they come after T1, T2, T3, and T4. It illustrates that the pitch contour and the register of T0 that comes after T1 are similar to those of T3, and the pitch contour and the register of T0 that follows T2 are similar to those of T4. The pitch contours of the other two tones are different from the full tones concerning pitch contour and register. Thus, we can compare the perception of double-duration T0 that comes after T1 and T2 with that of T3 and T4.

Thus T001 is inserted into the carrier sentence "ge3ge0 dou1 qu4 shang4ban1 le" (Even Gege went to work) in place of T3, and T002 is inserted into the carrier sentence "ge3ge0 dou1 qu4 shang4ban1 le" (Everybody went to work) in place of T4.

The subjects also listened to the four sentences with the original T1, T2, T3 and T4.

3.1.3. Procedure

During the perception test, every participant was provided with an answer sheet. Subjects listened to the synthesized syllables (words) with the modified tones and the original words. Then they were required to mark the degree of naturalness (natural or unnatural) on a 7-scale system, i.e. if it is perceived as the words listed in the answering sheet and it is natural, 7 is circled, but if the sound is different from the word listed in the answering sheet, then 0 is circled.

3.2. Results

Table 4 shows the mean score of naturalness for the modified words on a 7-scale system. T001 is the lengthened neutral tone originally after T1; T002 is the lengthened tone naturally after T2; T003 is originally after T3; and T004 is originally after T4.

### Table 4. Means of naturalness for the modified words.

<table>
<thead>
<tr>
<th>Modified tones (underlined)</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1+T001</td>
<td>6.74</td>
<td>0.93</td>
</tr>
<tr>
<td>T2+T002</td>
<td>6.74</td>
<td>0.65</td>
</tr>
<tr>
<td>T3+T003</td>
<td>6.58</td>
<td>1.02</td>
</tr>
<tr>
<td>T4+T004</td>
<td>6.95</td>
<td>0.23</td>
</tr>
<tr>
<td>T001+T0</td>
<td>6.05</td>
<td>1.39</td>
</tr>
<tr>
<td>T002+T0</td>
<td>6.47</td>
<td>1.02</td>
</tr>
</tbody>
</table>

Comparing the original neutral tones with the lengthened neutral tones, we found that there are no significant differences between them. In other words, there is no significant difference between the original T0 and the lengthened T0 that follows T1, F(1,36)=0.911, p=0.346; there is no significant difference between the original T0 and the lengthened T0 that follows T2, F(1,36)=1.756, p=0.193; neither is there any significant difference between the original T0 and the lengthened T0 that follows T3, F(1,36)=2.371, p=0.132; there is no significant difference between the original T0 and the lengthened T0 that follows T4, F(1,36)=1, p=0.324. These results indicate that native speakers naturally perceive the lengthened neutral tones as neutral tones. This reveals that duration does not play an important role during the perception of the neutral tones.
The comparison between the original word “Ge3ge” and the words that the first syllable is substituted by the lengthened T0 reveals that there is a significant difference between the normal Ge3ge and the double-duration T3, \( F(1,36)=7.628, p < 0.01 \). The comparison between the original word “Ge4ge” and the word that the first syllable is substituted by the lengthened T0 reveals that there is no significant difference between T4 and T0. \( F(1,36)=3.898, p = 0.056 \). The result indicates that the lengthened T0 that originally came after T2 sounds natural when it replaces T4. Meanwhile, the lengthened T0 that originally came after T1 sounds a little different from T3 when it replaces T3. However, the score of naturalness is over 6 for (Table 4) suggesting that this tone could still be perceived as T3, although it is not the same as the perception of the original T3 in the same word. On the other hand, the word Ge3ge (name) is not a commonly used word, possibly affecting perception. It shows that the neutral tone and full tones can be altered based on duration to some degree.

4. Discussion

The results of Experiment 1 and Experiment 2 indicate that pitch is dominant in the perception of the neutral tone while duration is not. In some cases, the neutral tone and full tones can be altered based on duration to some degree. In other words, some lengthened neutral tones can be categorized as full tones which are perceived mainly based on the pitch contour and register. Therefore, the tonal neutralization occurs in four phonetic environments, i.e., four full tones are indistinguishable after T1, T2, T3 and T4 respectively. Neutral tones could be categorized as four sandhi tones.

In Mandarin, the tone that carried by the second syllable changes according to the tone carried by the first syllable. The rule is similar to the sandhi rule in some northern Wu dialects, such as Shanghai dialect. Hirayama [7] suggested that tone sandhi (wide usage) in northern Wu dialects shares similar features with the neutral tone in Mandarin. In both cases, the tone carried by the second syllable changes while the tone carried by the first syllable does not. However, there are some differences between them.

First, tone sandhi applies for all bi-syllabic words in Wu dialects, while the neutral tone only occurs in some words, and we cannot predict where [12]. Second, previous studies indicated that shorter duration is another difference between the neutral tone and tone sandhi in Wu dialects. Nevertheless, the current study shows that duration does not play an important role during the perception of the neutral tone. This means the neutral tone is closer to tone sandhi in Wu dialects. Table 5 illustrates the tone sandhi of bi-syllabic and poly-syllabic words in Shanghai dialect.

### Table 5. Tone sandhi in Shanghai dialect.

<table>
<thead>
<tr>
<th>Bi-syllables</th>
<th>Tri-syllables</th>
<th>Five syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>55+21</td>
<td>55+33+21</td>
</tr>
<tr>
<td>5</td>
<td>33+44</td>
<td>33+55+21</td>
</tr>
<tr>
<td>6</td>
<td>22+44</td>
<td>22+55+21</td>
</tr>
<tr>
<td>7</td>
<td>33+44</td>
<td>33+55+21</td>
</tr>
<tr>
<td>8</td>
<td>22+23</td>
<td>22+22+23</td>
</tr>
</tbody>
</table>

Note: adapted from Shanghai Shiqu Fangyanzhi, 1988:24.

There are five tone categories in Shanghai dialect. In Table 5, the first column shows five types of tones carried by the first syllable in poly-syllabic words. The second, third and fourth columns show the pitch values of the tone sandhi in bi-syllabic words, tri-syllabic words and five syllabic words respectively. The sign “plus” indicates a boundary between two tones.

Tones carried by the second, third or fourth syllable change to other pitch contours depending on the tone carried by the first syllable. In the bi-syllabic words, the change of the second tone is similar to that of the neutral tone in Mandarin, which depends on the previous tones. If the previous tone is a high tone, the second tone is a mid-falling tone; if the previous tone is a mid-level tone, the second tone is a relatively high level tone; and if the previous tone is a low level tone, the second tone is a low rising tone. The second tone’s pitch contour, either a fall or a level, generally follows the first tone’s. In the tri-syllabic words, most of the last tones change to low falling tones, except the tone following the tone category 8.

This is the same as the neutral tone carried by the last syllables in tri-syllabic words. Chen and Xu [5] concluded that the neutral tone has a falling tendency according to the three consequent neutral tones. Actually, in Shanghai dialect, the last tone in the four consequent sandhi tones also has a falling tendency (Table 5). In this sense, the sandhi tones also have a falling tendency when the pitch of these tones depends on the previous full tones. The possible reason could be the low intensity.

Hirayama [7] claimed that the sandhi tone in the second syllable in disyllabic words in northern Wu dialects is a kind of neutral tone, which has low intensity, and they probably have the same origin. Zee’s finding (2008) that intensity is the covariant of pitch also supported the view that tone sandhi in Wu dialects shares the same feature, i.e., low intensity, with neutral tones in Mandarin. Therefore, the pitch of which the intensity is a covariant is dominant, and the neutral tones could be categorized in terms of pitch. As a result, four kinds of neutral tones occur in four phonetic environments in Mandarin, while five kinds of sandhi tones occur in Shanghai dialect.

The environment of the tonal neutralization for the neutral tone is its previous tone. Four full tones in Mandarin are indistinguishable when they follow the same full tone. Based on our observation, similar to sandhi tones in Shanghai dialect, the neutral tones generally follow the previous tones’ contours. In other words, if the previous tone is a non-low tone, the neutral tone is a falling tone; if the previous tone is a low tone, the neutral tone is a rising tone. Meanwhile, if there are two or three consequent neutral tones in Mandarin or sandhi tones in Wu, there is only a consequent falling contour across the tones or a falling contour with an initial rise if the ending point of the first tone is low.

Can we regard a neutral tone or a sandhi tone in Wu dialect as an extended part of the tone carried by the first syllable? If so, we can regard both the first full tone and the following tones as a whole tone. However, the initial and the finals are independent in a syllable carrying a neutral tone. Meanwhile, a syllable that carries a neutral tone or a sandhi tone still conveys meanings and it is written as a character. In other words, besides the neutral or sandhi tone that depends on the previous tone, other units in the syllable including segment, meaning and character are unchanged and independent. Therefore, we can only claim that the tone is dependent while the whole syllable is not. Some scholars [7] explained that the sandhi tone here is just to form stressed and unstressed syllables in a word instead of tonal distinction. Therefore, the second tone generally is weak and follows the previous tone.

On the other hand, the whole syllable that carries the neutral tone is the same as the previous syllable that carries full tone, although the feature of the neutral tone’s pitch is not stable. It could be regarded independent according to its
falling tendency. Wang [12] pointed out that it is common that the neutral tone has a falling tendency regardless the neutralization environments. The tonal pitch of T0 can change to be independent fall further, and the fall carried by this syllable can also appear in other contexts. For example, the base form of the tone carried by gui is T3. The tone is neutralized in meigui. so gui carries a high falling contour that is similar to T4. When gui occurs in guili, the tone carried by gui is not a neutral tone, but most people still produce it as T4 as well [12]. In the case, the neutral tone changes to an independent tone in another environment that is originally from its dependent pitch contour during tonal neutralization.

5. References