The Effects of Pitch Range and Duration on Tone Categorical Perception

Dazuo Wang$^{1,2}$, Gang Peng$^{1,3}$

$^1$ Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
$^2$ College of Chinese Language and Culture, Nankai University, Tianjin, China
$^3$ Language Engineering Laboratory, The Chinese University of Hong Kong, Hong Kong SAR

wandexmzuo@mail.nankai.edu.cn, gpengjack@gmail.com

Abstract

Previous research has mainly focused on the influence of language experience (cross-language or cross-dialect), and stimulus complexity (speech vs. non-speech) on the categorical perception of tones. This study further investigates the influence of pitch range and duration on tone categorical perception. Four speech tone continua differing in pitch range and duration were presented to native Mandarin listeners. Both identification and discrimination experiments were conducted. The results confirm that perception of speech tone continua for native Mandarin listeners is categorical, with discrimination peaks being well aligned with the identification boundaries. Pitch range was found to significantly influence the position of categorical boundary while duration only played a limited role in modulating the boundary position.

Index Terms: categorical perception, Mandarin, tone perception, pitch range, duration

1. Introduction

Categorical perception (CP) is the experience of perceptual invariance in sensory phenomena that can be varied along a continuum. Initially, for nearly 50 years, CP has been one of the most extensively studied phenomena in speech perception [e.g., 1, 2, 3].

For the segmental features of speech, it is generally thought that feature contrasts (e.g., voicing or place of articulation) of stop consonants are typically found to be perceived in a mostly categorical manner [1, 2], whereas contrasts involving vowel features are not, especially for those features related to the perception of single stable vowels [2, 3].

For the suprasegmental features of speech, there also exist a few pioneering studies of tone categorical perception. However, previous research has presented conflicting results with respect to the categorical nature of tone perception. By investigating Thai listeners’ perception of a continuum of three Thai level tones (low, mid and high), Abramson [4] reported that most participants could parse the continuum into three level tone categories. However, they exhibited uniformly good discrimination of stimuli across the continuum. Lack of discrimination peaks across the categorical boundary suggested that tone perception in Thai is not categorical. Wang [5], by using a speech continuum ranging from Mandarin Tone 1 to Tone 2, found that discrimination accuracy peaked around the tone category boundary. Wang suggested that native Mandarin listeners perceive tonal contrasts in a categorical manner. Because of the different language backgrounds of the subjects and different stimuli in the above two studies, it is not easy to compare them with each other. Using three tone continua (Taiwanese Tone 1 to 2; Taiwanese Tone 2 to 4; Taiwanese Tone 3 to 4), Hällé et al. [6] argued that the tone perception of Taiwanese listeners is in a quasi-categorical manner, with Taiwanese subjects showing higher degree of CP than that of French subjects. Despite the lack of lexical tonal contrasts in the French language, French listeners are not absolutely “deaf” to tonal variations.

To sum up, previous studies generally suggested that tone categorical perception is modulated by tone language and non-tone language experience. Whereas native tone language listeners perceive speech tone continuum categorically, non-native tone language listeners perceive the tone continuum more or less continuously.

Peng et al. [7] investigated not only the influence of tone language vs. non-tone language experience (German vs. Chinese), but also the influence of different tone inventories (Mandarin tones vs. Cantonese tones), on the categorical perception of pitch contour. They concluded that in the discrimination tasks, German listeners exhibit only psychophysical boundaries, whereas Chinese listeners exhibit linguistic boundaries, and these linguistic boundaries are further shaped by the different tone inventories.

Based on Peng et al.’s study, we further investigate whether pitch range and duration of tone continua affect the categorical tone perception. Gandour’s study [8] shows that the “average pitch height” and “direction” are the two dimensions underlying tone perception. As pitch range tightly correlates with the pitch height, we predict that different pitch range may affect tone perception. It is generally known that other suprasegmental characteristics, e.g., intensity profile and duration, are also correlated with tone perception [9, 10], while F0 is often taken as the primary acoustic cues of tone identity [10, 11]. In this study, we mainly focus on F0 and duration for the Mandarin tone perception, keeping other parameters constant.

2. Methods

2.1. Participants

Fourteen native Mandarin listeners (7 males, mean age = 24.8 yr, SD = 2.2) from SIAT (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences) were paid to participate in this experiment. Since different tones inventories of participants’ dialects and musical training may affect the perception results, we recruited participants who were from North China and got a high proficiency score (over 6, 1-9, where ‘1’ represents worst and ‘9’ represents best) in a Mandarin self-evaluation screen test. None of the participants had more than five years of formal musical training such as piano, violin, and none had any recent musical training within the past five years. No participant has difficulty in reading and hearing. Participants gave informed consent in compliance with a protocol approved by SIAT.

2.2. Speech materials

As shown in Figure 1, four tone continua differing in stimulus duration, i.e. 300 ms and 500 ms (not included in Figure 1),
and pitch range, i.e. 30 Hz and 50 Hz, were resynthesized from the syllable /i/ with high level tone uttered by a native Mandarin speaker. In each continuum, the Number 11 stimulus represents the syllable /i/ with the high level tone which means "衣 (clothes)" in Mandarin while the Number 1 stimulus represents the syllable /i/ with the high rising tone that means "姨 (aunt)" in Mandarin. All speech stimuli in the four continua were generated by applying the pitch-synchronous overlap and add (PSOLA) method implemented in Praat [12]. All speech stimuli were presented at 65dB.

Each continuum was resynthesized through the following procedures. Let us take Continuum 1 for instance: (1) We scaled the duration of the syllable /i/ to 300ms, and fixed the pitch contour to the level frequency 135 Hz. (2) Then, we reduced the number of pitch points to 3, with one at the onset position, one at the 60 ms position, and one at the offset position. (3) Dragging the above three pitch points accordingly, we synthesized the various stimuli for the tone continuum. The different starting frequencies for the Continuum 1 were determined by the formula \(105 \text{ Hz} + 3 \text{ Hz} \times \text{Stimulus position} + 3 \text{ Hz} \times (\text{Stimulus number}-1)\).

The pitch range of Continuum 1 is 30 Hz (from 105 Hz to 135 Hz), while the pitch range of Continuum 2 is 50 Hz (from 100 Hz to 150 Hz). The step size of Continuum 1 is 3 Hz, while the step size of Continuum 2 is 5 Hz. The Continuum 3 is a linearly lengthened version of Continuum 1, which is the same rising speech tone continuum used in [7]. And the Continuum 4 is a linearly lengthened version of Continuum 2.

2.3. Procedure

The experimental procedure and data analysis method have been adopted from [7, 13].

2.3.1. Identification test

In the identification test, participants listened to stimuli from each continuum presented in isolation. They were asked to

[Image of figure 1: Two 300-ms tone rising continua: (a) With pitch range of 30 Hz; (b) With pitch range of 50 Hz.]

2.4. Data analysis

Based on three essential characteristics of categorical perception: position of category boundary, width of category boundary and discrimination peak, we obtained individual measures for each participant so as to investigate the effects of pitch range and duration on identification and discrimination performance.

2.4.1. Identification scores

Given a particular stimulus, the identification score was defined as the percentage of responses with which participants identified that stimulus as being either "衣" (meaning clothes) or "姨" (meaning aunt). In a testing block, there were 2 repetitions of each of the 11 stimuli. There were 4 such testing blocks for each tone continuum. Prior to the testing blocks for each tone continuum, there was a practice block for familiarizing the participants with the test requirements in order to stabilize their performance in the testing blocks.

The same procedure was repeated for each of the four continua. We counterbalanced the order of four tone continua across participants.

2.3.2. Discrimination test

In the discrimination test, participants were run on an AX discrimination test. Stimuli were presented in pairs with a 500 ms inter-stimulus interval (ISI). For each continuum, there were 29 such pairs presented in a random order. Of these pairs, 18 consisted of two different stimuli separated by two steps, in either forward (1-3, 2-4, 3-5, 4-6, 5-7, 6-8, 7-9, 8-10, 9-11) or reverse order (1-3, 2-4, 3-5, 4-6, 5-7, 6-8, 7-9, 8-10, 9-11), and 11 consisted of the 11 stimuli each paired with itself (same pairs). There were 5 repetitions of each pair, resulting in 145 pairs in total being presented to each participant for each continuum. After hearing each pair, participants were asked to judge whether the two stimuli were the same or different, and to respond by pressing a key (‘v’ = “same” and ‘n’ = “different”). There was a feedback, the accuracy percentage shown after each block to encourage participants to concentrate on the test and do their best. There was also a practice block for each tone continuum before the testing blocks. The order of continuum presentation was also counterbalanced across participants.
BB trials. The score (or accuracy) $P$ for each comparison cohort was defined by:

$$P = P(\text{“S”}|\text{S}) \cdot P(\text{S}) + P(\text{“D”}|\text{D}) \cdot P(\text{D})$$

Where $P = P(\text{“S”}|\text{S})$ represents the percentage of “same” responses to all “same” pairs, namely the correct responses. Likewise, $P(\text{“D”}|\text{D})$ represents the percentage of “different” responses to all “different” pairs. $P(\text{S})$ and $P(\text{D})$ are the percentages of “same” and “different” pairs in each cohort, respectively.

3. **Results**

3.1. **Identification functions and discrimination curves**

Identification functions and discrimination curves are shown in Figure 2. The estimated boundary position and width in terms of stimulus step number, obtained by Probit analysis, are shown in Table 1.

Table 1. Categorical boundary position and width

<table>
<thead>
<tr>
<th>Continuum</th>
<th>Position</th>
<th>Corresponding starting frequency (Hz)</th>
<th>Width</th>
</tr>
</thead>
<tbody>
<tr>
<td>300 ms, 30 Hz</td>
<td>5.59</td>
<td>118.77</td>
<td>0.76</td>
</tr>
<tr>
<td>300 ms, 50 Hz</td>
<td>6.35</td>
<td>126.75</td>
<td>0.86</td>
</tr>
<tr>
<td>500 ms, 30 Hz</td>
<td>5.63</td>
<td>118.89</td>
<td>0.87</td>
</tr>
<tr>
<td>500 ms, 50 Hz</td>
<td>6.68</td>
<td>128.4</td>
<td>1.01</td>
</tr>
</tbody>
</table>

3.2. **Position of category boundary**

To investigate the impact of duration (300 ms, 500 ms) and pitch range (30 Hz, 50 Hz) on the category boundary position, we conducted a two-way ANOVA, with both variables as the within-subject factors.

The main effect of duration ($F(1, 13) = 3.22, p = 0.10$) was not significant, indicating that the mean boundary positions across the 300 ms, 500 ms duration continua were generally the same.

The main effect of pitch range ($F(1, 13) = 42.30, p < 0.0001$) was highly significant, indicating that the mean boundary positions were significantly different between the 30 Hz, 50 Hz pitch range continua. As shown in Table 1, with the same duration, the boundary positions in 30 Hz pitch range continua were smaller than that in 50 Hz pitch range continua. Corresponding starting frequency of the boundary position clearly shows that with the pitch range varying from 30 Hz to 50 Hz, the starting frequencies of category boundary positions also increase (from 118.77 Hz to 126 Hz for 300-ms continua, and 118.89 Hz to 128.4 Hz for 500-ms continua).

The duration $\times$ pitch range interaction effect was not significant, indicating that the effect of pitch range on the category position did not correlate with that of duration.

3.3. **Width of category boundary**

A similar statistical analysis was applied to boundary width. The two-way ANOVA revealed that there were no significant main effects for duration ($F(1, 13) = 0.26, p = 0.16$) and for pitch range ($F(1, 13) = 3.90, p = 0.07$) on the boundary width. It seems that the boundary widths in the four continua were generally the same.

3.4. **Discrimination peaks**

Discrimination curves for the four continua are also shown in Figure 2.

As for the 300 ms, 30 Hz continuum, the accuracy reached maximum, 78.6%, at pair 5-7, which was marginally significantly higher than 68.9% at pair 4-6 ($p = 0.051$), and significantly higher than 69.3% at pair 6-8 ($p < 0.05$).

For the 300 ms, 50 Hz continuum, the accuracy was highest at pair 6-8 where it reached 77.1%. There was a fuzzy maximum at pair 5-7, pair 6-8 and pair 7-9 (73.2%, 77.1%...
and 72.1% respectively), which was significantly higher than the discrimination accuracy reached in the adjacent pairs (63.6% at pair 4-6, p < 0.05, and 60.4% at pair 8-10, p < 0.01).

As for 500 ms 50 Hz continuum, the accuracy reached maximum at pair 7-9, 80.1%. Similarly, there was also a fuzzy maximum at pair 5-7, pair 6-8 and pair 7-9 (70.0%, 75.7% and 80.1% respectively), which was significantly higher than the discrimination accuracy reached in the adjacent pairs (55.7% at pair 4-6, p < 0.01, and 68.9% at pair 8-10, p < 0.01).

For the 500 ms 30 Hz continuum, a fuzzy maximum was located at pair 5-7 and pair 6-8 (74.6% and 68.9%, respectively), which was also marginally significantly higher than the discrimination accuracy reached in the adjacent pair 4-6 (65.7%, p = 0.052), and significantly higher than the discrimination accuracy reached in the adjacent pair 7-9 (59.3%, p < 0.05).

4. Discussion

4.1. Categorical perception of tones

As shown in Figure 2, the identification functions and discrimination curves of the four continua generally exhibit the classical patterns of category perception for native Mandarin listeners, according to the well-known categorical perception characteristics [15]: (1) A sharp boundary between two categories in the identification function; (2) accuracy peaks at the category boundary, but is at or near chance level within category in the discrimination curves; (3) the predictability of discrimination from identification. The findings in present study are consistent with previous cross-language studies [5, 6, 7, 13] that the tone perception by native tone language listeners is categorical for speech tone continuum.

4.2. The significant effect of pitch range on the category boundary position

In our daily speech communication, the pitch range of a syllable may indeed stretch or compress on different occasions. Chao [16, 17] argued that the actual pitch movement of Mandarin speech is the algebraic sum of tone and intonation. If a rising intonation coincides with a rising tone, the pitch of the rising tone will be higher than usual. If a falling intonation coincides with a rising tone, the rising tone will rise less or start at a lower register. For example:

1. 他姓王? (Is his surname Wang?)
2. 他姓王。 (His surname is Wang.)

In sentence (1), “王” is a rising tone, which coincides with a rising intonation, so it is higher than usual. Whereas in sentence (2), the second clause is with a falling intonation, so the rising tone of “王” here rises less. Therefore the pitch range is dynamic, sometimes larger or sometimes smaller in a certain range.

In the current study, there was a highly significant effect of pitch range on the category boundary position. As shown in Table 1, it seems that, with the pitch range varying from 30 Hz to 50 Hz, the onset F0 of stimuli at category boundary positions also increase.

Just as seen in Figure 3, the pitch range is just like a “spring” which can be stretched or compressed. There is a small point just like the onset F0 of category boundary position. The small point rises while the spring is being stretched (the upper bound of the pitch range being stretched more than the bottom).

Pike [18] thought that there is a linguistic relativity: There are sounds which are approximately the “same” in physical terms, but can be sharply “different” in terms of their relation to the context and the system in which they occur. He further claimed that linguistic relativity for the interpretation of segments refer not only to the structural, functional positions in sequences, but also to the specific individual sounds which precede and follow them.

![Figure 3. Schematic explanation of category boundary shift with respect to the expansion of pitch range.](image3)

He supposed that a certain language signifies contrastive pitch heights which carry different meanings. A linguistically high syllable is not necessary with high absolute F0, but relatively high in comparison to its neighboring syllables. As shown in Figure 4, three weights suspended between springs, illustrating three level tones in a tone language. They are not quite at a state of rest, so there is some slight variation in their position. In addition to this slight variation, however there may be much larger variation in accordance with various speaking modes. For example, a quiet speaking mood may bring the tones closer together, excitement may raise all of them up, and sudden force at one point may displace one weight more than the others.

![Figure 4. Three weights, not quite at a state of rest, suspended between springs, illustrate three level tones in a tone language (adapted from Figure 1(a) in [18]).](image4)

In our study, the influence of the supporting springs on the weights is just like that of intonation on tones. We can draw a picture as Figure 5.

![Figure 5. Another schematic explanation of category boundary shift with respect to the expansion of pitch range.](image5)
Vividly, we can see that in spite of absolute changes, the tone system as a system remains the same. They are still separated from each other by relative height. The relative height, similar to the pitch range, is dynamic in a certain range. With the pitch range stretched while with the low bound fixed or slightly changed, the category boundary position rises in the relative pitch space. We further infer that the category boundary position would decrease with the pitch compressed.

The results indicate that the tone category boundary is a range rather than an absolute line. Within the range, there are ambiguous stimuli that neither sounds like “姨” or “姨” to the participants. Such ambiguous stimuli may appear in daily communication, and participants need more time or more contextual information to recognize them. Whereas, the ending points stimuli are typical “姨” or “姨”, participants can easily perceive them as two corresponding categories.

5. Conclusion

In this study, we investigate the effects of pitch range and duration on Mandarin tone categorical perception. As for native Mandarin listeners, the speech tone perception is categorical, with discrimination peaks being generally well aligned with the categorical boundary positions. Pitch range significantly affects the categorical boundary position. With the pitch range varying from 30 Hz to 50 Hz, the boundary position rises. However, it seems that duration only plays a limited role in modulating categorical boundary position.

6. Acknowledgements

The work described in this paper was partially supported by grants from National Science Foundation of China (NSFC: 11074267, 61135003). We are thankful to all the subjects at SIAT who participated in this time-consuming and attention-demanding experiments. We thank the members of the Language Engineering Laboratory of The Chinese University of Hong Kong for many helpful discussions, and Quansheng Xia and Ye Liu of Language Cognition Group at SIAT for their constructive help.

7. References