The perception study of Mandarin Tone 1 and Tone 4 by Hong Kong Cantonese speakers: The pitch effects

Mai Zhang Yi Liu

Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University, Hong Kong
mai.zhang@connect.polyu.hk yi.liu@polyu.edu.hk

ABSTRACT

Mandarin has four lexical tones and Hong Kong Cantonese has six lexical tones. The pitch pattern of T1 and T4 in Mandarin is high level and high falling. According to the results of previous research, the pitch boundary of T1 and T4 by native speakers of Mandarin is categorical. Using the synthesized disyllable real words as stimuli, this paper aims to examine the identification and discrimination of Mandarin T1 and T4 by 24 native speakers of Hong Kong Cantonese. The finding of this study can provide reference to Cantonese studies and language teaching in Hong Kong.

Keywords: Mandarin, T1, T4, perception, Hong Kong Cantonese speakers.

1. INTRODUCTION

In the tonal system of Mandarin. T1 is high level and T4 is high falling. T1 and T4 contrast by pitch height at the end of the syllable. With regard to the speech perceptual studies on Mandarin tones, substantial investigations have been done by scholars. Wang [7] approved the existence of perception categorisation between T1 and Tone 2 in Mandarin by using synthesised monosyllable continuum. Hall et al. [2] and Xu et al. [9] observed the effects of pitch contours on the categorical perception of tones. Wu et al. [8] reported the effects of linguistic experiences on the perceptual assimilation of lexical tones.

In terms of Mandarin tonal perception study by tonal and non-tonal speakers, So & Best [6], Li & Shuai [3] used monosyllables to study the perception of Mandarin T1 and T4 by native speakers and non-native speakers with different language background. The two allophonic variations (55 and 53) of T1 in Hong Kong Cantonese may have effect on the perception of Mandarin T1 and T4. The aim of this paper is to use disyllabic words to study how native speakers of Hong Kong Cantonese perceive high level and high falling tones in Mandarin.

2. EXPERIMENT DESIGN

Most previous studies on Mandarin tone perception dealt with synthesized monosyllabic words. Xue & Shi [10] used disyllabic Mandarin words and conducted the perception test to discover the pitch boundary of T1 and T4 by native speakers. The current work focuses on the perception of Mandarin T1 and T4 by native speakers of Hong Kong Cantonese with synthesized disyllable real words. Identification and discrimination tasks are adopted in the experiment with the application of E-prime. We will compare the data of this experiment with Xue & Shi's results.

2.1. Stimulus

Through a five scale familiarity test that involved 42 university students, 8 pairs of disyllable words were selected as stimuli in this experiment. These disyllabic Mandarin words contrast by T1 and T4 either in first syllable or second syllable, and have different tone syllables as reference. The recording of the words embedded in a sentence was produced by a male native speaker of Mandarin with Cool Edit software. The sampling rate is 11025Hz. The pitch range of all recorded words is between 9 to 19 semitones by the transformation between Hz to semitone. Applying Praat to synthesize and normalize the words recorded, we have following referential syllables: 19-19 semitones, 160ms (T1); 14-19 semitones, 160ms (Tone 2); 9-9 semitones, 140ms (Tone 3); 140ms, 19-14 semitones (T4 at first syllables), 19-9 semitones (T4 at second syllables). As target syllables of 160ms and 19-19 semitones, 11 stimulus materials are synthesised with 50Hz as a reference to convert Hz to semitone and one semitone difference at endpoint.

2.2. Participants

Twenty four (12 female and 12 male) university students of native Hong Kong Cantonese speakers participated in the experiment. Their average duration of Putonghua study was 8 years. The average age of the participants was 21 years with

SD of 1.91. And all participants were right handed without visual disability.

2.3. Tasks

Both identification task and discrimination task are conducted with the application of E-prime. The participants were asked to take a trial test before undertaking the formal test. There were a total 88 stimuli in the identification test. The recording of sounds played randomly for each participant. Two words appeared on the screen each time. There were two groups in this task. In group one the target words of T1 appeared on the left side of the screen and in group two the target words of T4 appeared on the left side of the screen. When hearing a sound, participants were required to make selections by pressing the designated computer button. The participants pressed button F to choose the words on the left side of the screen, and button J for the words on the right side.

The discrimination task adopted Pollock and Pisoni's AX method with 72 stimulus materials. Two sounds were played with the differences of 2 semitones in pitch and 500ms interval. Participants were asked to judge whether the two sounds were the same and press the designated computer button. There were two types of response on screen: same-different and different-same. And there were two kinds of stimuli order: low-high and high-low. The criterion for stimuli grouping in the discrimination task is displayed in the following table.

Table 1. The criterion for stimuli grouping in discrimination task.

types	stimuli order	response on screen
1	low-high	different-same
2	high-low	different-same
3	low-high	same-different
4	high-low	same-different

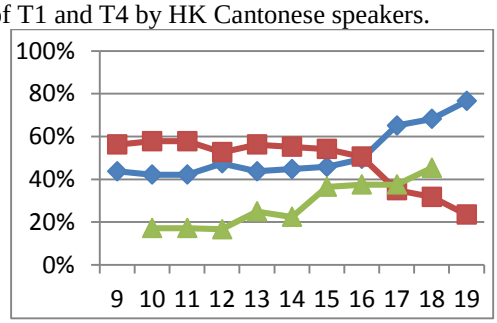
3. RESULTS AND ANALYSIS

Based on the E-prime data, we calculated the identification rate. The perceptual data is displayed by x axis of pitch semitones and y axis of identification rate. By analyzing the data generated from E-prime, we obtained the boundary location of high level and high falling pitch, pitch of turning point of identification curve, discrimination peak, discrimination rate and etc. ANOVA was adopted to carry out further analysis.

3.1. Boundary and discrimination

According to Liberman et al. [4] and Repp et al. [5], the categorical boundary of speech perceptual curves has three features: the appearance of steep rising and falling of identification curve, the peak of discrimination curve and the matching point (categorical boundary) of intersection of identification curves and discrimination peak.

Figure 1: The identification and discrimination of T1 and T4 by HK Cantonese speakers.



The width of curve refers to the distance between minimum and maximum value of identification curves. The obtained results (Figure 1) show that there is no matching point between the intersection of identification curves and discrimination peak. The identification curves after the intersection points are smoother than that before the points. The width of identification curves after the boundary is bigger than that before the boundary. The discrimination curve is not smooth with the peak value of over 40%. The intersection point of identification curves is located between 15 and 16 semitones with a discrimination rate of 50%. Compared with the results by Xue & Shi [10], we found the perceptual boundary pitch of identification curves by native speakers of Mandarin is higher than that of native speakers of Hong Kong Cantonese.

3.2. First and second syllables

The table below lists the maximum of identification curves for T1 and T4 in both first and second syllable groups. The maximum identification of T1 as target syllables in the first syllabic words group was 88%, whereas the maximum identification of T1 as target syllables in the second syllabic words group was 96%. The maximum identification of T4 as target syllables in both first syllabic words group and second syllabic words group was 83%. In the first syllabic words group the identification curves are smoother after the intersection points than the ones before the points. Moreover, the width of curve in the first syllabic group is bigger that in the second syllabic group. The result also indicated that the maximum value of identification curves in T1 was bigger than that of T4 and the maximum value of identification curves in the second syllabic group was bigger than that in the first syllabic group.

Table 2. The maximum value of identification curves of T1 and T4 in disyllable word groups.

	stimuli	maximum of T1	stimuli	maximum of T4
F	T1T1	88%	T1T4	83%
	T1T2	54%	T4T2	71%
	T1T3	88%	T4T3	46%
	T1T4	67%	T4T4	58%
S	T1T1	79%	T1T4	83%
	T2T1	96%	T2T4	38%
	T3T1	88%	T3T4	71%
	T4T1	71%	T4T4	75%

With regard to the discrimination analysis, the discrimination rate for both curves is close to 18 semitones. The discrimination curve of the first syllabic group is smooth. However, the discrimination rate is quite low with a discrimination peak of 41%.The data also indicates that the discrimination peak of target syllables in the first syllabic group is from 29% to 50%, whereas the discrimination peak of target syllables in the second syllabic group is from 63% to 71%. By comparison we found that the discrimination peak of target syllables in the second syllabic group is higher than the target syllables in the first syllabic group. That means the categorical perception of target syllables in the second syllabic group is better that of target syllables in the first syllabic group. The latter feature obtained from this study is consistent with Xue & Shi's [10] reports on native speakers.

3.3. Gender differences

In both gender groups the identification curves before the intersection point are not smooth. And the width of two identification curves in the female group is bigger than that of the male group.

For both male and female groups the discrimination rates are below 50%. For the male group the peak of discrimination appears around 18 semitones with a discrimination rate of 44%. Similarly for the female group the peak of discrimination appears around 18 semitones with a discrimination rate of 44%. Both curves show no falling after the peak. In both groups there is no pitch matching between the turning point of the identification curve and discrimination peak.

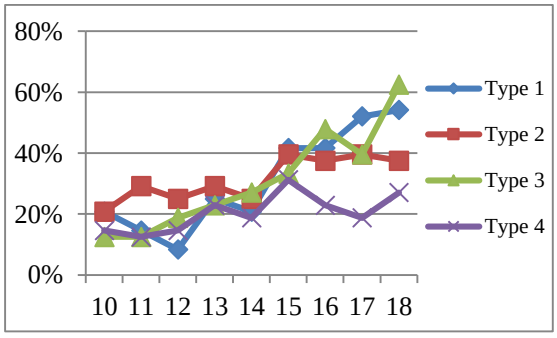
The pitch difference between the location of the intersection point and discrimination peak is 2 semitones in the male group, whereas the pitch difference between the location of the identification boundary and discrimination peak is 1.9 semitones in the female group. Applying ANOVA to conduct comparative analysis of discrimination curve of male and female speakers, it is found that only point 1 ($F=7.1772$, $P<0.05$*) and point 2 ($F=8.0476$, $P<0.05$*) show significant differences on gender.

In the tonal perception model of native speakers, the pitch of the intersection of two identification curves matches with the discrimination peak [10]. This phenomenon provides evidence of the categorical boundary for tonal perception. As the participants in this experiment are non-native speakers, their identification curves such as the male group, have more than one intersection with low discrimination peak. The perceptual pattern of male speakers manifests some degree of non-categorisation perception of high level and high falling pitch by male participants. By comparison the perceptual pattern of females shows the intersection of identification curves, which is the proven perceptual boundary of high level and high falling pitch. To sum up, the perceptual identification and discrimination of Mandarin high level and high falling tones for female participants is better than that for male participants.

3.4. Effects of stimuli order and word order

As mentioned previously, there are four types of discrimination tasks with different orders of stimuli and word orders. Figure 2 depicts the discrimination curves of four types of stimuli order and word order.

Figure 2: The mean discrimination of four types of stimuli order and word order.



Compared with type 2 with the peak of 15/17 semitones and type 4 with the peak of 15 semitones, the discrimination peaks of type 1 (54%) and type 3 (63%) with the peak over 50% are higher than that of type 2 (40%) and type 4 (31%). In another words, the discrimination peak of "low-high" stimuli order was higher than "high-low".

A Two way ANOVA indicated no significant interaction between stimuli order and word order [$F(1, 32) = 0.322$, $p = 0.574$]. Statistical results also showed that there is no significant effect on other types of order grouping. However, One way ANOVA found that word order had significant

effect on the high-low stimuli order [F (1, 16) = 12.308, $p < 0.005$]. That means the tonal perception of high-low stimuli order was affected by word order. Therefore we can draw the conclusion that the word order has an effect on the tonal perception of Mandarin T1 and T4. As affected by both the factors of word order and stimuli order, the discrimination rate of type 4 was the lowest among all the order types.

4. DISCUSSION AND CONCLUSIONS

Different from pervious works, the current project uses stimuli of disyllable words with different pitch at first syllables and second syllables as target units to investigate how Hong Kong Cantonese speakers perceive Mandarin high level and high falling tones. Based on the analysis above, we obtain the following conclusions. The pitch locations of identification boundary in the group of the second syllabic words are higher than that in first syllabic words group. Similarly, the discrimination rate in the second syllabic words is bigger than that in the first syllabic words group. Gender also plays an important role in tonal pitch categorical perception. The female participants are better at acquiring and perceiving Mandarin high level tone and falling tone than male participants. The results also indicate that influence of experiment on the perceptual categorization depends on a variety of word order factors. The effect of word order on Mandarin tonal perception is more obvious than stimuli order.

In Hong Kong Cantonese T1 has two variations: high level (55) and high falling (53). Due to the influence of their native tonal system, Hong Kong Cantonese speakers usually produce tonal errors when speaking Mandarin. Zeng [11] points out that they often pronounce the high level tone in Mandarin as high falling tone or vice versa. Because of the existence of two phonemes in T1 of Hong Kong Cantonese, the native speakers of Hong Kong Cantonese cannot differentiate the Mandarin high level tone and high falling tone as easily as native speakers of Mandarin. The results of this study can also be explained by Best's [1] Perceptual Assimilation Model. Although the participants learn Putonghua for 8 years, the learners cannot distinguish two lexical tones in the target language. Affected by the two phonemes in Hong Kong Cantonese, the Putonghua learners of Hong Kong Cantonese regard T1 in Hong Kong Cantonese as T1 or T4 in Mandarin. This leads to the difficulty in perceiving Mandarin T1 and T4 by Hong Kong Cantonese speakers. The results obtained from this project support So & Best's conclusion. In order to improve the categorical perception of T1 and T4 in Mandarin, adaptive perceptual training methods proposed by Zhang [12] can be adopted in the teaching process as an effective teaching method. The finding of this study can provide reference to language acquisition research, Cantonese studies and language teaching at Hong Kong.

5. ACKNOWLEDGEMENT

This study is supported by the project G-UC64 from Hong Kong Polytechnic University.

6. REFERENCES

[1] Best, C. T. 1995. A direct-realist view of cross-language speech perceptions. In W. Strange (Eds.) *Speech perception and linguistic experience: theoretical and methodological issues.* Timonium, MD: York Press, 229–273.

[2] Hall, P., Chang, Y. H. & Best, C. T. 2004. Identification and Discrimination of Mandarin Chinese Tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*. 32, 395–421.

[3] Li, B. & Shuai, L. 2011. Effects of native language on perception of level and falling tones. *Proc. 17th ICPhS* Hong Kong, 1202–1205.

[4] Liberman, A. M., Harris, K. S., Hoffman, H. & Griffith, B. C. 1957. The discrimination of speech sounds within and across phonemic boundaries. *Journal of experimental psychology*. 53, 358–368.

[5] Repp B. H., Healy, A. F. & Crowder, R. G. 1979. Categories and context in the perception of isolated steady state Vowels. *Journal of experimental psychology:* human perception and performance. 5, 129–145.

[6] So, C. K., & Best, C. T. 2010. Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech.* 53, 273-293.

[7] Wang, S. 1976. Language change. Annals of the New York academy of sciences 28, 61–72.

[8] Wu, X. H., Munro, M. J. & Wang, Y. 2011. The effects of linguistic experience on perceptual assimilation of lexical tone. *Proc. 17th ICPhS* Hong Kong, 2169–2173.

[9] Xu, Y., Gandour, J. T. & Francis, A. L. 2006. Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *J. Acoust. Soc. Am.* 120, 1063–1074.

[10] Xue, X. & Shi, F. 2012. Putonghua yinping he qusheng zhijian de tinggan fenjie. *Experimental Linguistics.* 1, 69–75.

[11] Zeng, Z. F. 1996. Difficult points in learning Putonghua in Hong Kong. *Dialect.* 2, 110–112.

[12] Zhang, L. J. 2009. Adaptive perceptual training and the categorical perception of aspirated/unaspirated contrast by Japanese learners of Chinese. *Chinese Teaching in the World*. 4, 560–566.