EFFECTS OF LANGUAGE TRAINING ON MANDARIN LEXICAL TONE PERCEPTION BY JAPANESE SPEAKERS

Wu Qi

University of Tsukuba
wuqi920524@yahoo.co.jp

ABSTRACT

In this study, we investigated the perception of Mandarin tones with different registers and syllable durations among Japanese speakers with different kinds of language training. The results showed that language training type and registers affected tone perception: the longer the language training, the closer the tone perception was to that of native Mandarin speakers. Register and syllable duration also affected the perception of Mandarin tones: the perception was more difficult in the low register, particularly for less experienced learners. Less experienced learners performed better on the shortest syllable duration. Also, for both learners and native speakers of Mandarin, the threshold values of ΔF0 (|onset – offset|) necessary to identify Tone 4 were larger than that of Tone 2.

Keywords: tone, years of language training, register, syllable duration, perception.

1. INTRODUCTION

There are four phonemic tones in Mandarin. Tone 1 (T1 hereafter) has a high-level pitch, Tone 2 (T2 hereafter) has a high-rising pitch, Tone 3 (T3 hereafter) has a low-dipping pitch, and Tone 4 (T4 hereafter) has a high-falling pitch [3]. According to [5] and [6], native Mandarin listeners identify tones by employing multiple acoustic cues. Furthermore, perceptual studies have suggested that F0 and F0 contour are the fundamental perceptual cues of Mandarin tones, e.g. [7].

How Mandarin learners identify tones is also a topic of interest. Previous studies have conducted experiments on native speakers of English, e.g. [4], but only a few studies have been conducted on Japanese Mandarin learners whose native language is Japanese which is non-tonal, e.g. [8] or [9]. In this study, we investigated the effects of three factors (length of language training, register, and syllable duration) on Japanese learners' tone perception and compared them with that of native speakers.

2. METHOD

2.1. Stimuli

Speech materials were developed from Mandarin bā (T1), produced by a single female native speaker. The native speaker pronounced it with three other tones at her comfortable pitch range. It was recorded through Praat [2] with 44.1 kHz, 16-bit digitization. The female voice was chosen as the original stimulus so that stimuli with a wide range of pitch can be created without the loss of naturalness.

Fig. 1 is a schematic representation of synthesized stimuli for this study where the onset and offset for each stimulus were varied systematically. All stimuli were synthesized based on stimuli0 in Fig. 1 whose onset and offset were 280.5Hz. VocalShifter LE [1] was used for the synthesis.

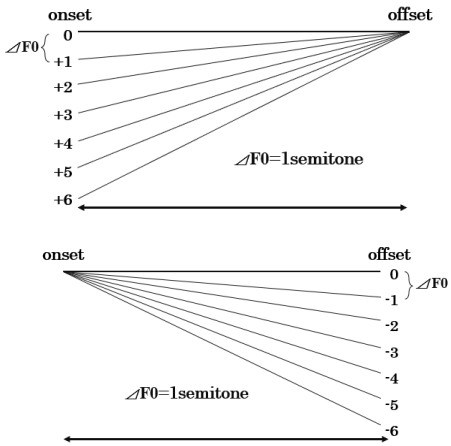

Figure 1: Schematic diagram of synthesized F0 patterns

The onset F0 and stimulus durations were manipulated subsequently to create stimulus sets with three F0 ranges (“register” hereafter), and three durations. F0 range was set to 25-31 semitone in the high register, 12-18 semitone in the medium register, and 5-11 semitone in the low register relative to 100Hz. Three durations, 350ms, 550ms, and 750ms, were used. In each register and syllable duration, thirteen stimuli with stepwise rising or falling pitch contours as seen in Fig.1 were included, resulting in 117 stimuli in total.

2.2. Participants

16 native speakers (NS) and three groups of 26 Japanese speakers participated in this study (Table 1). The first group (G1) learned Mandarin for 2 months to 1 year, the second group (G2) for 1.5 to 3 years, and the third group (G3) for 3.5 to 6 years. Participants learned Mandarin in regular undergraduate classes, and some continued their studies after graduation. We classified the learners based on the length of training rather than proficiency test scores because the test scores were not available for many of the participants. All participants were reported to have normal hearing.

Group	Length of Training	Gender	Age
G1	2 months - 1 year	2M, 12F	20.71
G2	1.5 years - 3 years	2M, 6F	22.38
G3	3.5 years - 6 years	2M, 2F	22.25
NS	Native speaker	6M, 10F	25

Table 1: Participants.

2.3. Procedures

This study used a four-alternative forced-choice task. Each participant listened to a stimulus and selected a tone they perceived, from "bā", "bá", "bǎ" and "bà" written in pinyin (the Romanized spelling system of Chinese characters with tone symbols indicated by diacritics). A calculation task was inserted after each response to avoid biases from the previous stimulus, the stimuli were presented in three blocks by syllable durations (350ms, 550ms, 750ms). The order of stimuli was randomized within each block. Each participant listened to the stimuli at his/her comfortable level over headphones on a computer. The participants were allowed to take a break after each block. All participants completed the experiment in 30 to 45 minutes.

3. RESULTS

Fig. 2 shows the responses in ratio for each stimulus by groups, tallied across syllable durations and registers. Lines connecting symbols represent responses to the four tones in Mandarin: red for T1, green for T2, yellow for T3, and blue for T4. The judgment curves by NS showed sharp boundaries between T1 and T2 as well as T1 and T4. Near 100% responses were observed for T1, T2, and T4 at some points along the stimulus continuum, indicating clear categorical perception. Responses for T3 did not show the clear peak seen in T1, T2, and T4, and the stimuli were rarely identified as T3 by NS. This indicates the stimulus continuum in this study did not include a stimulus that can be perceived clearly as T3 by native Mandarin speakers[1].

The judgment curves by G1 were less steep and the response ratio for T1, T2, and T4 did not reach near-100 percent anywhere along the continuum. It is of interest that G1 had T3 responses rather often compared with other groups. The judgment curves of G3 have distinct areas with high identification rates for T1, T2, and T4, similar to those of NS.

The maximum responses along the continuum (maximum identification rates, hereafter) for T1, T2, and T4 (i.e., the peaks in Fig. 2) increased with years of language training. For T1 and T4, the maximum identification rates were near 100% in G3 and NS but were lower in G1. Notice that for T2, the maximum identification rate occurred at S+2 in G1 but S+6 in G3 and NS.

3.1. Response of T1, T2 and T4

Table 2 shows the maximum identification rates for T1, T2, and T4 by registers and groups. Data were tallied across durations.

For T1, the maximum identification rates of G1 learners were 52% to 57%, which were substantively lower than those of the other groups in all registers.

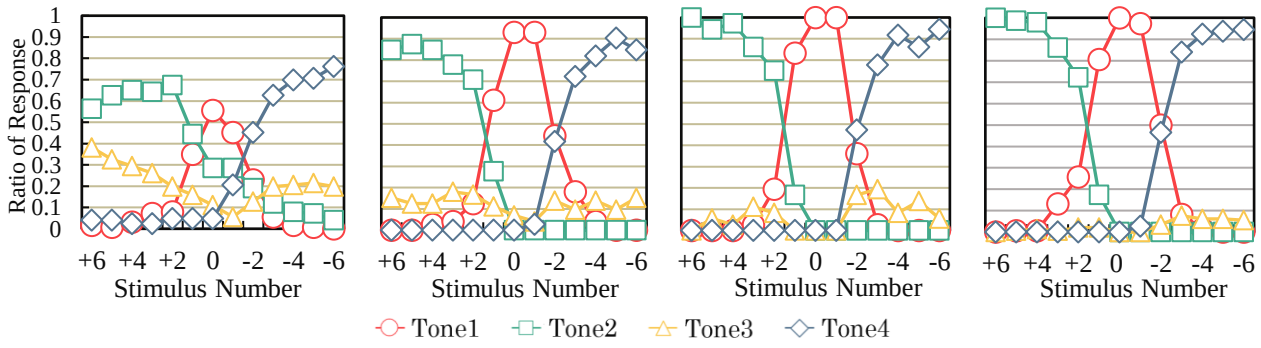

Figure 2: Responses in ratio for each tone and stimulus. From left to right, each figure depicts results of G1, G2, G3 and NS.

The maximum identification rates for G2, G3, and NS were near 100% across registers.

For T2, the maximum identification rates were 100% across registers in G3 and NS. However, G2 showed a lower maximum identification rate at the low register compared to the high and medium registers. In G1, the maximum identification rate declined as the register became lower. All groups showed a decline in maximum identification rates at the low register for T4.

Table 3 summarizes the maximum identification rates of T1, T2, and T4 by groups and by syllable durations. For T1, T2, and T4, the effects of language training were observed in all durations. No apparent effect of syllable duration was observed in any groups for T1. For T2 and T4, the highest maximum identification rates in G1 and G2 were achieved at T=350ms, with substantively lower rates at longer durations.

3.2. Perceptual boundaries between lexical tones

Phonological perceptual boundaries were calculated for each participant in each register and syllable duration. Results from T1 responses were used. G1 data were not included in this analysis because they did not demonstrate sufficient categorical perception.

First, the perceptual judgment curve was approximated by a logistic function (see equation (1) below) for each participant in each register and duration. Second, the locations where the curve crossed the 50% identification threshold were identified. Two locations, representing T1T2 and T1T4 boundaries, were identified for each curve and defined as the phonological perceptual boundary values. The horizontal axis in Fig. 3 shows the phonological perceptual boundary values and the left and right panels show the T1T2 and T1T4 boundaries, respectively. HR, MR, and LR indicate high, medium, and low registers, respectively. Each symbol represents the mean phonological value for the specified group, register, and syllable duration. The error bars depict one standard deviation around the mean for each group, register, and duration.

$$f(x) = \frac{1}{1+e^{ax+b}} \quad (1)$$

The T1T2 boundaries are closer to S0 than the T1T4 boundaries. This means the amount of descent required for the perception of T4 is greater than the rise required for the perception of T2.

For each of the T1T2 and T1T4 boundaries, a three-way ANOVA was conducted with group, register, and syllable duration as explanatory variables and perceptual boundaries as the objective variable.

For T1T4 boundaries, the main effects of group, register, and syllable duration were not significant and there were no significant interactions. For T1T2 boundaries, the main effect of the register ($F(2, 50)=17.854$, $p<.05$) and of syllable duration ($F(2, 50)=3.827$, $p<.05$) were significant. There was a significant interaction between syllable duration and register ($p<.05$). Post-hoc analysis revealed that the

Response / Register	T1				T2				T4			
	G1	G2	G3	NS	G1	G2	G3	NS	G1	G2	G3	NS
High register	57	100	100	100	86	100	100	100	81	100	100	100
Medium register	57	100	100	100	74	100	100	100	76	96	100	100
Low register	52	92	100	100	52	71	100	100	71	75	83	83

Table 2: The maximum identification rates for T1, T2, and T4 by groups and registers.

Response / Duration	T1				T2				T4			
	G1	G2	G3	NS	G1	G2	G3	NS	G1	G2	G3	NS
T=350ms	57	96	100	100	74	96	100	100	81	100	92	96
T=550ms	55	96	100	100	64	83	100	100	76	88	100	96
T=750ms	55	96	100	100	69	88	100	100	76	83	92	94

Table 3: The maximum response rates for T1, T2, and T4 by durations.

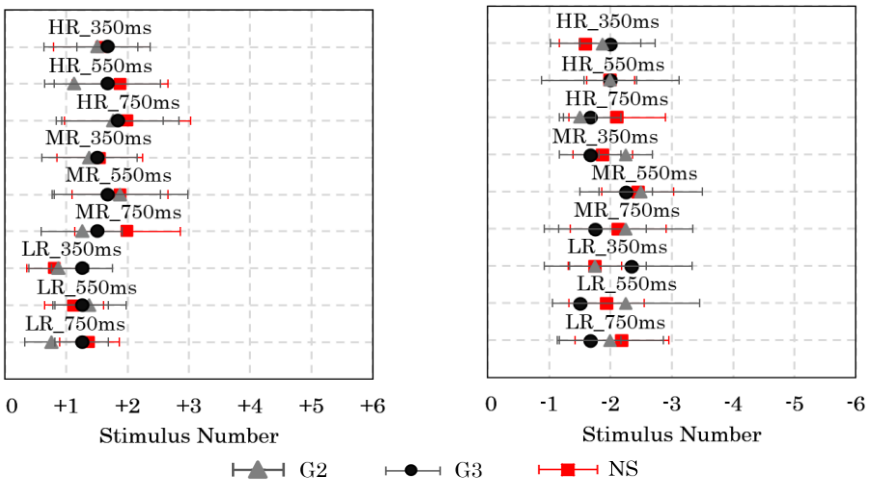

Figure 3: Relationship of perceptual boundary and stimulus number from participants.

differences between T=350ms vs. T=550ms and T=350ms vs. T=750ms were significant (p<.05). T1 T2 boundary was closest to S0 in T=350ms. The differences between the low vs. high and the low vs. medium register were significant (p<.05): the T1T2 boundary was closest to S0 in the low register. In other words, the lower the register, the smaller the rise required for a stimulus to be perceived as T2, and the shorter the syllable duration, the smaller the rise required to be perceived as T2.

4. DISCUSSION

The results of the present study suggest that the longer the language training, the closer the judgment curves were to those of native Mandarin speakers. Learners with language training lasting more than 3.5 years (G3) may have formed a system of categorical perception of Mandarin tones similar to that of native speakers.

Voice register is also likely to have effects on the perception of these tones: learners with a shorter period of language training, such as G2 and G1, did not identify T2 as well as G3 and NS in the low register. Syllable durations also appeared to have effects, suggesting that perception was easier with short syllable duration for listeners with less language training.

Examination of perceptual boundaries revealed that the threshold values of ΔF0 (|onset – offset|) necessary for identifying T4 were larger than that for T2. In other words, listeners were more sensitive to the rising tone than to the falling tone. It was also found that the location of the T1T4 boundary was affected by registers and by syllable durations. Taken together, the results provide quantitative evidence of the acquisition of tone perception over the course of language learning among Japanese speakers.

Although stimuli that can be perceived as typical T3 (a low-dipping tone) were not present in this study, some participants, especially G1 listeners, perceived T3 occasionally. The criteria for judging T3 is an issue for future research.

5. ACKNOWLEDGEMENTS

I am grateful to Prof. Zhu Chunyue and Yasunori Takahashi for their insightful comments. Additionally, I extend my gratitude to the anonymous reviewers for their detailed and valuable comments. This research was supported by JSPS KAKENHI Grant Number JP22K13161.

6. REFERENCES

[1] AckieSound. 2016. VocalShifter LE (Version 3.1) [Software].

[2] Boersma, P., & Weenink, D. 2018. Praat (Version 6.0.42) [Software].

[3] Chao, Y. R. 1948. Mandarin primer. Cambridge, MA: Harvard University Press.

[4] Chun, D. M., Jiang, Y., Meyr, J., & Yang, R. 2015. Acquisition of L2 Mandarin Chinese tones with learner-created tone visualizations. Journal of Second Language Pronunciation, 1(1), 86–114.

[5] Gandour, J. 1984. Tone dissimilarity judgments by Chinese listeners. Journal of Chinese Linguistics, 12, 235–261.

[6] Massaro, D. W., Cohen, M. M., & Tseng, C. 1985. The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. Journal of Chinese Linguistics, 13, 267–290.

[7] Wayland, R. P., & Guion, S. G. 2004. Training English and Chinese listeners to perceive Thai tones: A preliminary report. Language Learning, 54(4), 681–712.

[8] Zhang Linjun. 2010. Riben liuxuesheng hanyu shengdiao de fanchouhua zhijue. *Yuyan Jiaoxue yu Yanjiu* (in Chinese.). 3, 9–15.

[9] Zhu Hong. 2013. Nihonjin gakushusha ni yoru Chugokugo seichou no shuutoku no kenkyu (in Japanese.). Ph.D. Dissertation. Tohoku University.

[1] This may be because the stimuli in this experiment, as seen in Fig. 1, did not contain the low-dipping pitch pattern that is a key feature of T3.