DISCRIMINATION OF CANTONESE TONES BY MANDARIN, ENGLISH AND FRENCH SPEAKERS

Zhen Qin and Peggy Pik-Ki Mok

Department of Linguistics and Modern Languages, Chinese University of Hong Kong
qinzhenquentin@gmail.com

ABSTRACT

This study tests the effects of L1 experience and psychoacoustic similarity of stimuli on the naïve perception of Cantonese tones by Mandarin, English and French speakers. AX discrimination tasks of speech and non-speech tones were conducted. Results showed that the subjects performed differently between the speech and non-speech tasks. While the three L1 groups shared some confusable tone pairs due to acoustically similarity, they differed in specific pairs under the influence of L1 prosodic system.

Keywords: tone perception, L1 influence, psychoacoustic similarity.

1. INTRODUCTION

In cross-linguistic perceptual studies concerning naïve listeners, listener's native language (L1) often exerts an influence on their perceptual performance [1]. In addition, language-independent factor, psychoacoustic similarity, also plays an important role in speech perception and affects all listeners similarly [2].

The present study focuses on the perception of Cantonese tones by speakers of both tone and non-tone languages. The objective is to explore how and to what extent L1 prosodic system together with psychoacoustic similarity of stimuli affect the way naïve listeners attend to Cantonese tones.

Cantonese is a tone language, in which tones are lexically defined. There are six contrastive lexical tones (T) according to [3]: T1 [55] High Level; T2 [25] High Rising; T3 [33] Mid Level; T4 [21] Low Falling; T5 [23] Low Rising; T6 [22] Low Level. As shown in Figure 1, T1 stands out from the other tones in terms of pitch height. The mid level tones, T3, are further apart from T1 than the low level tone, T6. The tonal space in the lower pitch range is very crowded. The two rising tones T2 and T5 share the starting point. They only differ in the magnitude of rising pitch movement. Additionally, T4-T6 and T5-T6 differ only in the final part. T4 falls slightly while T5 rises slightly towards the end. Taken together, the psychoacoustic similarities between these tones may cause confusion for all listeners.

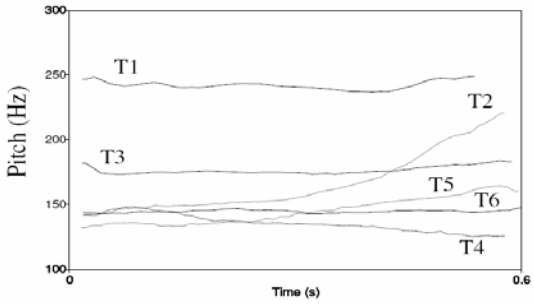

Figure 1. F0 traces of the six Cantonese tones.

Most of the previous studies focused on the perceptual differences between speakers of tone and non-tone language in the discrimination and identification of tonal contrasts [4, 5]. As far as we know, the perceptual differences between speakers of two non-tone languages are not investigated yet. Thus, speakers of two non-tone languages (English and French) together with speakers of one tone language (Mandarin) are involved in this study. Although English and French do not have lexical tones, they use pitch differently: English is a lexical stress language and French is a language without lexical prosody. Pitch variation is used in the syllable-level to contrast lexical stress in English, whereas French speakers may show a low sensitivity to pitch variation in the syllable-level. It is hypothesized that the Mandarin group can distinguish Cantonese tones better than the other two groups due to their linguistic experience of native tones. However, whether

and how the prosodic differences between English and French would result in perceptual differences awaits investigation.

2. METHOD

2.1. Subjects

There were 12 Mandarin (2 M, 10 F), 10 English (7 M, 3 F) and 10 French (3 M, 7 F) native speakers in this study. They were all university students, aged between 18 and 26. All were naïve listeners without specific Cantonese learning experience. They had no or only limited music training and they reported no speech or hearing impairments.

2.2. Stimuli

Two syllables, /jau/ and /se/, each carrying six Cantonese tones, were used as test stimuli. One female native speaker of Hong Kong Cantonese was recorded reading the target syllables carrying six tones in a carrier phrase ---*我读_字* "I read the word___." three times. The target syllables were excised and in total twelve tone stimuli (2 syllables × 6 tones) were chosen.

Besides the tones in natural speech, non-speech tones were used as control stimuli. The stimuli were pure tones with simple harmonics, synthesized from the six Cantonese tones with the syllable /jau/ produced by a different female native speaker. The pure tones have similar F0 profiles and duration to the six tones carried by /jau/ and /se/.

2.3. Procedures

AX discrimination tasks of both speech and non-speech stimuli were conducted. First, all the possible pairings of the six tones with each linguistic syllable, including 6 AA and 15 AB pairs for each monosyllable, were used and presented randomly to the subjects. The presentation order was counter-balanced in the AB pairs. There were altogether 72 tokens (15 AB pairs × 2 syllables × 2 orders + 6 AA pairs × 2 syllables) with 60 AB pairs and 12 AA pairs[1]. The inter-stimulus interval (ISI) was 500 milliseconds (ms). The presentation of the stimuli was controlled by the software DMDX with a laptop computer.

After participating in the discrimination task of speech tones, the same subjects took part in an AX discrimination task of pure tones. All the possible pairings of the six pure tones, including 6 AA and 15 AB pairs, were presented randomly. Each of AB pairs was presented four times with presentation order counter-balanced and each AA pair was presented ten times. There were altogether 120 tokens (15 AB pairs × 4 times + 6 AA pairs ×10 times) with 60 AB pairs and 60 AA pairs. The ISI was 500 ms. The whole process was controlled by E-prime 2.0 Professional with a desktop computer.

The procedures of the speech and non-speech tasks were the same. The stimuli were presented to subjects through a stereo headphone with the volume adjusted to a comfortable level in a quiet room. The subjects were told that they would hear pairs of sounds from a certain language. They were required to discriminate two sounds in each pair as fast and as accurately as possible by pressing a button referring to "same" on the left side using their left index finger and a button referring to "different" on the right side using their right index finger. Missing responses were excluded from analysis. No feedback was given. A short practice was given before each task. The whole experiment lasted approximately 30 minutes.

3. RESULTS

Both error percentage (EP) and reaction time (RT) for correct pairs were recorded. Since all the participants made very few errors for the AA pairs in both speech and non-speech tasks, only the results of the AB pairs are analyzed.

3.1. Speech vs. Non-speech tasks

The EP and RT (collapsed across presentation order and across the two monosyllables) of the speech and non-speech tasks are shown in Figure 2. The results of the two tasks are compared in order to test the L1 influence on the subjects' performance in the speech task.

Two Repeated-Measures ANOVA tests were conducted for EP and RT separately with L1 Group (Mandarin, English, and French) as the

between-subjects factor and Task (speech task vs. non-speech task) as the within-subjects factor. In terms of EP, the results revealed significant main effects for L1 groups [F (2, 29) =4.47, p=0.02], and tasks [F (1, 29) =55.62, p<0.001]. Crucially, the interaction effect [F (2, 29) =3.75, p=0.036] was found. In terms of RT, the results found significant main effects for L1 groups [F (2, 29) =7.44, p=0.002] and tasks [F (1, 29) =72.43, p<0.001] but no interaction [F (2, 29) =2.09, p=0.14>0.05].

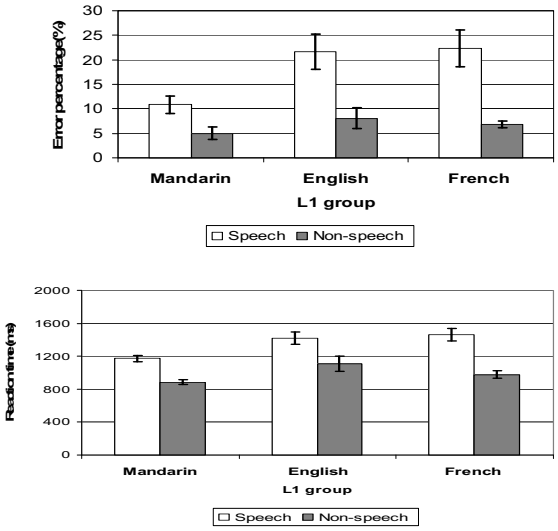

Figure2. Error percentage (top panel) and reaction time of correct pairs (bottom panel) in speech and non-speech tasks by different L1 groups.

Post-hoc comparisons (Bonferroni) found that the Mandarin group did better than the English and French groups in both EP (p=0.037; p=0.026) and RT (p=0.026; p=0.007) in the speech task. In the non-speech task, the effect of L1 group was only found for RT, which is mainly due to the difference between the Mandarin and English groups (p=0.023). No significant difference was revealed between the Mandarin and French groups, and between the English and French groups.

To sum up, the overall difference between the speech and non-speech tasks demonstrated that the subjects did not hear the speech and non-speech tones in the same way. Linguistic experience was much reduced in the non-speech task. On the contrary, the difference among the L1 groups in the speech task can be attributed to the subjects' L1 experience. Mandarin speakers performed much better than the English and French speakers.

3.2. The effect of individual tone pairs

The perceptual performance of each L1 group was examined in detail with respect to individual tone pairs in the speech task. Figure 3 shows the EP and RT of the 15 AB pairs by different L1 groups. Two Repeated-measures ANOVA tests were conducted on the EP and RT with L1 Group (3 levels) as the between-subjects factor and Tone Pair (15 levels) as the within-subjects factor.

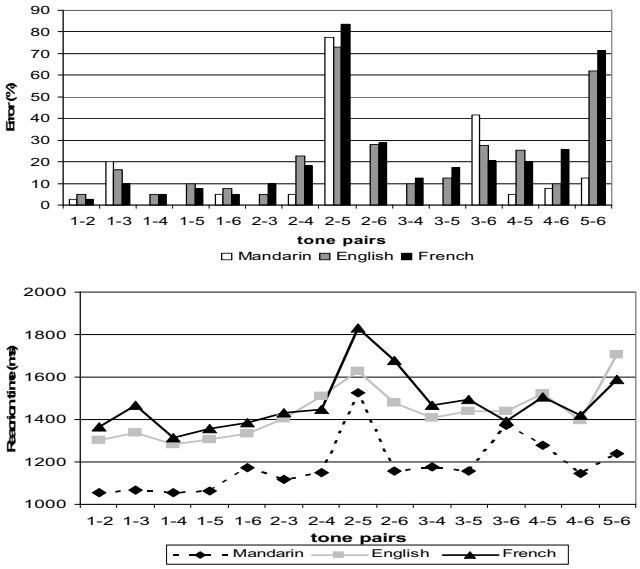

Figure3. Error Percentage (top panel) and reaction time (bottom panel) of the 15 tone pairs in three L1 groups.

In terms of EP, the results yielded significant effects of L1 groups [F (2, 29) =4.95, p=0.014], tone pairs [F (5.8, 168) =41.24, p<0.001], and crucially, the interaction effect [F (11.6, 168) =3.24, p=0.05]. In terms of RT, the results revealed significant effects of L1 groups [F (2, 29) =6.65, p=0.004 <0.05] and tone pairs [F (7.4, 215) =8.69, p<0.001], but no interaction [F (14.8, 215) =1.06, p=0.396>0.05].

The obvious patterns in Figure 3 demonstrated that for each L1 group, fewer errors were found for the pairs with T1 than other tone pairs. The T2-T5 pair had the highest EP and longest RT. While the level tone pairs such as T3-T6 and T1-T3 were difficult for the Mandarin subjects (with high EP), the tone pairs

including contour tones such as T5-T6 were quite difficult for the English and French subjects, resulting in the second highest EP.

4. DISCUSSION

Both psychoacoustic similarity of stimuli and L1 experience were found to affect the naïve perception of Cantonese tones.

With regard to psychoacoustic aspects, the three groups shared similarities in discriminating some pairs due to the psychoacoustic similarity of these tones in the speech task. First, all the subjects found the pairs with T1 easier to distinguish than the other tone pairs because T1 is well separately from the other tones in the acoustic space (see Figure 1). They found the pair of T2-T5 the most confusable because T2 and T5 are acoustically similar and only differ in the magnitude of the final rising movement. Second, among the level tones, T3-T6 had a higher EP and a longer RT than T1-3, T1-6 for all the L1 groups. The shorter acoustic distance between T3 and T6 than between T1 and T3 contributes to the relative difficulty of this pair among the level tone pairs for all the L1 groups.

Regarding L1 influence, the performance of the three L1 groups in the speech task was different from the non-speech task. A clear different performance between the Mandarin group and the other groups was found in the speech task whereas most of the differences among the three L1 groups in the non-speech task were not statistically significant. Therefore, we conclude that L1 experiences did exert a much greater influence on the speech task than on the non-speech task.

More important, a different perceptual pattern between the Mandarin group and the two L1 groups was found in the speech task. First, the Mandarin group did much better in terms of EP and RT than the other L1 groups. The best performance of the Mandarin group should be explained by their linguistic experience with native tones. Second, the three groups had different performance on specific pairs. The Mandarin group found the level tones difficult to distinguish. Owing to only one level tone in the tone inventory, the Mandarin speakers lose the sensitivity to differences of the level tones, which are within-category differences for them. No significant difference was found between the English and French groups. The results are opposed to the prediction in [4] that French speakers may outperform English speakers because French prosody does not use pitch in any word-level constraint (i.e. lexical stress) as English does. Owing to the lack of lexical tones in the native prosodic system, the English and French groups perceived tones mainly relying on psychoacoustic aspects of the stimuli. They could hardly discriminate some similar tones such as T5 and T6, but they distinguished the level tones better than the Mandarin speakers.

5. REFERENCES

[1] Best, C. T. (1995). A direct realistic view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Baltimore: York Press.

[2] Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, *75*, 1866-1878.

[3] Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. New York: Mouton de Greyter.

[4] Halle, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32, 395-421.

[5] So, C. K. (2010). Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences. *Language and Speech*, 53(2), 273-293.

[1] The unbalance of the AB and AA pairs may have induced bias to the "different" responses resulting in more errors for the AA pairs. However, only very few errors of the AA pairs were found for each L1 group. Additionally, this study focuses on the results of AB pairs. Therefore, the unbalanced design did not appear to have affected the results adversely.