

PERCEPTION OF CANTONESE TONES BY MANDARIN, ENGLISH AND FRENCH SPEAKERS

Zhen Qin & Peggy Pik-Ki Mok

Department of Linguistics and Modern Languages, Chinese University of Hong Kong, Hong Kong
qinzenquentin@gmail.com; peggymok@cuhk.edu.hk

ABSTRACT

This study tests the influence from L1 experience and psychoacoustic similarity on the naïve perception of Cantonese tones. Three groups of subjects, Mandarin, English and French listeners, participated in an AX discrimination task and a dissimilarity rating task. The discrimination results showed that while the three L1 groups shared some confusable tone pairs which are acoustically similar, they differed in specific pairs under their L1 influence. The rating task found that the different perceptual performances are revealed in terms of weight assigned to different dimensions such as pitch direction and height by the three L1 groups.

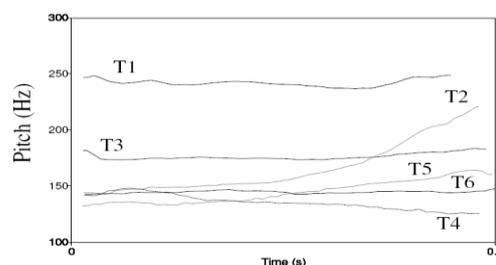
Keywords: perception, Cantonese tones, L1 experience, psychoacoustic similarity

1. INTRODUCTION

In cross-linguistic perceptual studies concerning non-native sounds perceived by naïve listeners, listener's native language (L1) often exerts an influence on their perceptual performance. In addition, language-independent factor, such as psychoacoustic similarity, also plays an important role in speech perception and affects listeners in a universal way.

Different distribution of attention to specific dimensions is proposed to explain the perceptual differences across different L1 speakers. Speakers of tone and non-tone language were found to differ in the weightings assigned to dimensions, especially pitch height and direction. Speakers of non-tone languages were more sensitive to pitch height than direction, whereas speakers of tone language placed more emphasis on pitch direction [2, 3]. 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 the L1 prosodic systems together with psychoacoustic similarity affect the way naïve listeners attend to non-native tones.

Figure 1: F0 traces of the six Cantonese tones.



Cantonese is a tone language, in which tones are lexically defined. There are six contrastive lexical tones (T) according to [1]: 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 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 (T2-T5, T3-T6, T4-T6 and T5-T6) may cause confusion for all listeners.

Three L1 groups are involved in this study. Mandarin is a lexical tone language while English and French are non-tone languages. English is a lexical stress language and French is a language without lexical prosody. Pitch is introduced into the phonological representation of the target languages at different levels. The prosodic differences between English and French should be noted. First, pitch variation is used in the syllable-level to contrast lexical stress in English, whereas French lacks lexical prosody. French listeners show a low sensitivity to pitch variation in the syllable-level [4]. Second, English appear to allow for more intonation contour variations than French [6]. It is hypothesized that the Mandarin group can distinguish Cantonese tones better than the other

two groups because of 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. METHODS

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 students in the Chinese University of Hong Kong, aged from 18 to 26. All were naïve listeners without any specific Cantonese learning experience. All the listeners had no or only limited music training and they reported no speech or hearing impairments.

2.2. Stimuli

Three syllables [ji], [se], [jau], 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 eighteen tone stimuli (3 syllables × 6 tones) were chosen.

2.3. Procedures

Two tasks were used in the experiment. First, a forced-choice AX tone discrimination task, using the syllables of [se] and [jau], was performed. All the possible pairings of the six tones with each syllable, including 6 AA and 15 AB pairs, were presented randomly. The presentation order was counter-balanced in the AB pairs. There were altogether 72 tokens (2 syllables × 6 tones (1st sound) × 6 tones (2nd sound)). The inter-stimulus interval (ISI) was 500 milliseconds (ms). The presentation of the stimuli was controlled by the software DMDX.

Second, a dissimilarity rating task, using the syllable [ji] with 36 tokens (6 tones (1st sound) × 6 tones (2nd sound)), was done. The subjects were required to judge the dissimilarities between each pair of stimulus by circling a number on a 9-point dissimilarity scale (“1” = no differences, “5” = “medium differences”, “9” = “extreme differences”). The ISI was 500 ms.

Practice was given before each task. The subjects were required to judge as fast and as accurately as possible.

3. RESULTS

3.1. Discrimination

The data of both reaction time (RT) and error percentage (EP) were collected, but only EP data are reported here due to page limit. Since all the participants made very few errors for the AA pairs, only the AB pairs were analyzed. Repeated-measures ANOVA was conducted on the EP (collapsed across presentation order and across two syllables) with L1 group as between-subject variable (3 levels) and tone pair (15 levels) as within-subject variable.

Figure 2: Error Percentage of the 15 tone pairs across different L1 groups.

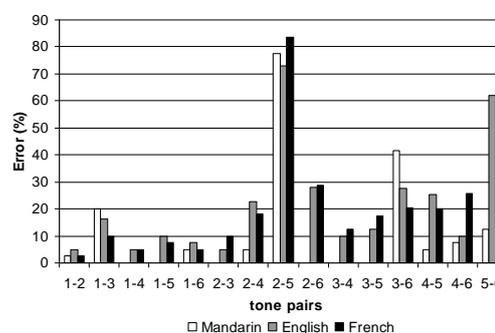
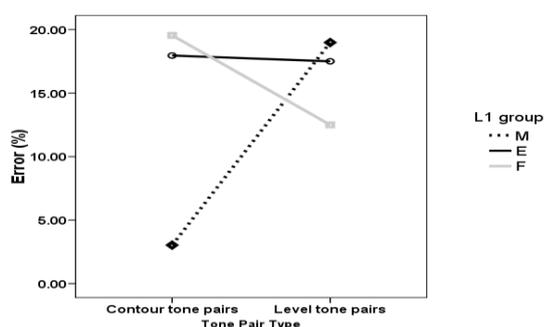


Figure 2 shows the EP in different L1 groups. The mean EP for Mandarin, English and French subjects are 10.88%, 21.67% and 22.33% respectively. Results revealed significant main effects of L1 group [$F(2,29)=4.95, p=0.014$], tone pair [$F(1,29)=41.24, p<0.001$], and interaction effect [$F(2,29)=3.24, p=0.05$]. The Mandarin group performed significantly better than the English ($p=0.044$) and French ($p=0.03$) groups, but no difference was found between the English and French groups ($p=1.00$). Within group pairwise comparisons of tone pairs revealed that for each L1 group, fewer errors were found for the pairs with T1 than other pairs; the T2-T5 pair had the highest EP. The Mandarin group had difficulty in distinguishing the level tones such as T3-T6. The English and French groups found the pair of T5-T6 hard to discriminate, resulting in the second highest EP. The English group distinguished T4-T6 and T5-T6 better than the French group.

The T2-T5 pair is confusable for all the L1 groups. In order to investigate the effects of L1, the other 14 tone pairs were divided into two categories: Level (T1-T3, T1-T6, and T3-T6) vs. Contour (the other 11 pairs). As Figure 3 shows, the Mandarin group made much fewer errors for the contour pairs than for the level pairs, whereas

the pattern was reverse for the English and French groups, although there is only a small difference for the English group. The results were consistent with [3, 2] in that Mandarin speakers were sensitive to pitch direction and both English and French speakers only placed emphasis on height. While level and contour pairs appeared to be equally difficult for the English group, the French group discriminated level pairs better than contour pairs.

Figure 3: Error Percentage (%) of contour vs. level tone pairs across different L1 groups.



In sum, the results of EP showed that all the subjects found the pairs with T1 easy to distinguish and the pair of T2-T5 most confusable. Additionally, the three L1 groups differ on some specific pairs. While the level tone pairs such as T3-T6 and T1-T3 were difficult for the Mandarin subjects, the contour tone pairs such as T5-T6 were quite difficult for the English and French subjects. The overall performance of the Mandarin group was much better than the other groups; the English group performed slightly better than the French group, but the difference was not statistically significant.

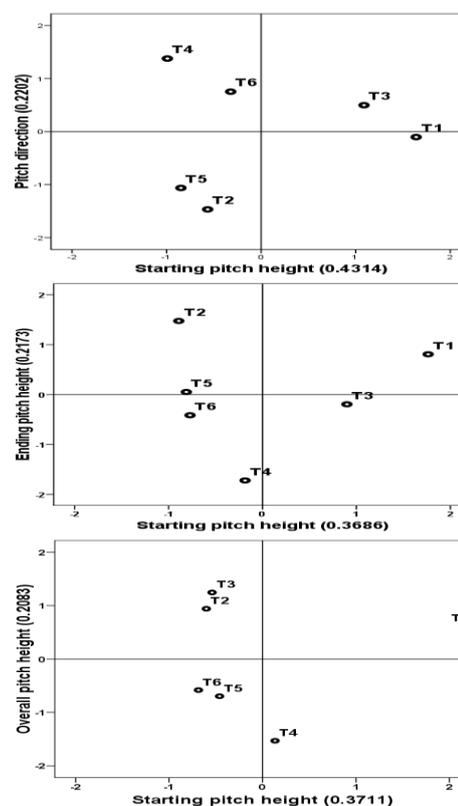
3.2. Dissimilarity rating

The results of the dissimilarity rating task were submitted to a Multidimensional scaling (MDS) analysis using the INDSCAL model. A two-dimension solution for each group was calculated and the results are shown in Figure 4.

In terms of dimension interpretation, for the Mandarin group, D1 is labeled as “Starting pitch height” with T1, T3 (high) on one side and other tones (low) on the other side. D2 corresponds nicely to “Pitch direction” with the rising (T2, T5) and falling (T4) tones positioned at two ends and the three level tones falling between them. For the English group, the tones along D1 are separated in a similar way to the Mandarin group, so the same label is given. D2 is interpreted as “Ending pitch

height”, as T1, T2 with the highest ending pitch and T4 with the lowest are distributed toward the two ends. For the French group, with T1 separated from the other tones, D1 could be also labeled as “Starting pitch height” despite the fact that the tonal distribution along D1 for the French group is less clear-cut than that for the other two groups. D2 is interpreted as “Overall pitch height”, as the Cantonese tones of high (yin) (T1, T2, T3) and low (yang) (T4, T5, T6) registers are roughly separated along this dimension.

Figure 4: MDS two-dimension solutions for Mandarin (top panel; stress=0.218, RSQ=0.652), English (middle panel; stress=0.242, RSQ=0.586) and French group (bottom panel; stress=0.313, RSQ=0.51).



In terms of tonal distribution, T1 is often separated from the other tones. While T2 and T5 are close in the perceptual space of the Mandarin group, the two tones are also close along D1 for the English and French groups. T5 and T6 are much closer in the perceptual space of the English and French groups than the Mandarin group. The three level tones, T1, T3, and T6, are clearly closer in the perceptual space of the Mandarin group than the other two groups. The three L1 groups, especially the Mandarin and English groups, separated the tones in a similar way along D1 but

in different ways along D2. In general, while the six tones are distributed relatively separately for the Mandarin and English groups, they are distributed closely for the French group.

The results of the rating task are partially consistent with the findings in the discrimination task. First, given “Starting pitch height” shared by the three L1 groups, the Mandarin group placed more weight on pitch direction whereas the English and French groups were only sensitive to height. Second, the perceptual distance of the six tones is closely related to the discrimination performance. For instance, T2-T5 was difficult for the Mandarin group to distinguish and the two tones were close in the perceptual space of the Mandarin speakers. Likewise, a high EP of T5-T6 was found for the English and French groups and the two tones had a short distance for both groups. Although the confusion of T2-T5 for the English and French groups and that of T3-T6 for the Mandarin group in the discrimination task did not agree completely with the corresponding perceptual distance in the rating task, the two tones in each pair is close along either D1 or D2. The discrepancies between the two tasks should be attributed to task specific factors such as stimuli used.

4. DISCUSSION

With regard to psychoacoustic aspects, it is not surprising that all the subjects found the pairs with T1 easy and the pair of T2-T5 most difficult to discriminate. T1 is well separated from other tones in the acoustic space; T2 and T5 are acoustically similar and only differ in the magnitude of the final rising movement. Even native Cantonese speakers found this pair confusable [5]. Moreover, the confusion of T3-T6 for the Mandarin group and T5-T6 for the English and French groups is partially due to the acoustic similarity between these tones. Among the three groups, starting pitch height was relied on by all the groups and appeared to be acoustically salient for listeners to separate the six tones.

Regarding L1 influence, the three groups had different performances on specific pairs although they shared some confusable pairs due to acoustic similarity. The Mandarin listeners found the level tones difficult to distinguish comparing with the other groups; The English and French groups could hardly discriminate some contour tone pairs such as T5-T6. Due to no tonal contrast of level tones

in Mandarin tonal inventory, the Mandarin group is less sensitive to pitch height than direction and have difficulty in distinguishing the level tones. In contrast, as tone does not have a phonemic status in English and French, the two groups are not sensitive to pitch direction, a linguistic cue. Thus, the English and French groups found the contour tones difficult.

The English and French groups did distribute attention to different cues because of their L1 prosodic differences. For English, pitch variation is used in stress, so the English listeners can detect the pitch height variations in syllable-level. Furthermore, boundary tones, a higher or lower tone, are aligned with the initial and final edges of words or phrases in the English intonation system. Taken together, the English listeners should have sensitivity to pitch variations in the starting and ending parts. For French, although pitch is also used in intonation contours, the lack of lexical prosody and the position-fixed accent at the final edge result in that unlike the English listeners, the French listeners were not sensitive enough to ending pitch height. As the six Cantonese tones differ mainly at the ending part, the English listeners, because of their sensitivity to ending pitch height, discriminated some pairs such as T4-T6 and T5-T6 better than the French listeners.

On the whole, the study found that psychoacoustic similarity and L1 experience affected the perception of Cantonese tones by speakers of tone and non-tone languages.

5. REFERENCES

- [1] Bauer, R.S., Benedict, P.K. 1997. *Modern Cantonese Phonology*. New York: Mouton de Greyter.
- [2] Francis, A.L., Ciocca, V., Ma, L., Fenn, K. 2008. Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics* 36, 268-294.
- [3] Gandour, J.T. 1983. Tone perception in far Eastern languages. *Journal of Phonetics* 11, 149-175.
- [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] Mok, P., Wong, P. 2010. Perception of the merging tones in Hong Kong Cantonese: preliminary data on monosyllables. *Proceedings of Speech Prosody 2010* Chicago, 100916, 1-4.
- [6] Vaissi ère, J. 2002. Cross-linguistic prosodic transcription: French versus English. In Volslkaya, N.B., Svetozarova, N.D., Skrelin, P.A. (eds.), *Problemy i Metody Eksperimental'no-foneticheskikh Issledovanij. In Honour of the 70th Anniversary of Prof.L.V. Bondarko*. St. Petersburg: St.-Petersburg State University, 147-164.