The perception of speech and non-speech tones by tone and non-tone language listeners

Zhen Qin, Peggy Mok

Department of Linguistics and Modern Languages, Chinese University of Hong Kong

qinzhenquentin@gmail.com, pegggymok@cuhk.edu.hk

Abstract

This study tested how linguistic experience and psychoacoustic factors affect tonal perception of Mandarin, English and French speakers. AX discrimination tasks of speech and non-speech tones were conducted. Results showed that the subjects performed differently in the speech and non-speech tasks. In the speech task, while the three L1 groups shared some confusable tone pairs due to their acoustical similarity, they differed in specific pairs under the influence of L1 linguistic experience. In the non-speech task, however, the three L1 groups did not have different error patterns of individual pairs. In short, both L1 experience and psychoacoustic similarity of stimuli were found to have an impact on the perception of non-native tones.

Index Terms: Tone perception, Linguistic experience, Psychoacoustic similarity, Cross-Linguistic study.

1. Introduction

1.1. Psychoacoustic factors vs. Linguistic experience

In cross-linguistic perceptual studies concerning naïve listeners, listeners' native language (L1) often exerts an influence on their perceptual performance [1]. A language-independent factor, psychoacoustic similarity of stimuli, also plays an important role and affects all listeners similarly [2]. While L1 experience affects the perception of speech sounds, acoustic factors should play a more dominant role in non-speech perception. The present study includes speech and non-speech tones in order to explore how L1 experience together with psychoacoustic similarity of stimuli affect naïve listeners' perceptual performance.

1.2. Speech vs. Non-Speech Tones

Although prior studies [3] found that linguistic experience extended its influence to the processing of non-speech tones, previous study suggested that listeners differed in perceiving speech and non-speech tones. A clear difference between speech and non-speech tasks is that L1 influence was attested in speech task but not in non-speech task in [4]. Likewise, categorical perception of Cantonese tones was attested for native speakers in perceiving speech tones rather than non-speech tones in [5].

Both natural stimuli of Cantonese tones and non-speech tones with analogous pitch movement were used in this study in order to test the L1 influence, which may differentiate the speech task from the non-speech task. If a clear difference between the two tasks is found, it is taken as evidence that L1 influence prevails in the speech task only.

1.3. Pitch in different languages

Cantonese has six contrastive lexical tones (T) according to [6]: 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 tone, T3, is 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.

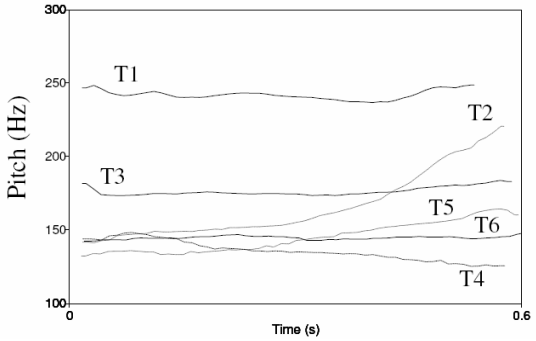

Figure 1. *F0 traces of the six Cantonese tones.*

Most of the previous studies focused on the perceptual differences between speakers of tone and non-tone languages in the discrimination and identification of tonal contrasts [7, 8]. As far as we know, the perceptual differences between speakers of two non-tone languages are not empirically investigated yet. Thus, speakers of two non-tone languages (English and French) together with speakers of one tone language (Mandarin) are involved as subjects.

Although being closely related to Cantonese, Mandarin tonal inventory has only one level tone (55). The others are contour tones: a dipping tone (214), a rising tone (35), and a falling tone (51).

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 [7]. Pitch variation is used in the syllable-level to contrast lexical stress in English, whereas French has a rising pitch pattern on the last syllable and does not contrast meaning in stress. Thus, compared with English speakers, French speakers may show a lower 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. 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 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 the six tones in a carrier phrase ---*我读_字*"I read the word___". The target syllables were excised and in total twelve tone stimuli (2 syllables × 6 tones) were chosen.

Besides the Cantonese tones in natural speech, non-speech tones were used as control stimuli. The stimuli were pure tones with one frequency component, resynthesized from the six Cantonese tones in Praat with the syllable /jau/ produced by another female native speaker. The pure tones have similar F0 contours 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 by using two types of materials: AA pairs (pairs with same tone) and AB pairs (pairs with different tones).

2.3.1. Speech task

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. Each AB pair was presented two times with presentation order counter-balanced. 60 tokens of AB pairs (15 AB pairs × 2 syllables × 2 orders) and 12 tokens of AA pairs (6 AA pairs× 2 syllables) were used[1]. The 72 tokens in total were grouped into 7 blocks with 10 tokens in each block and 12 pairs in the last block.

The inter-stimulus interval (ISI) was 500 milliseconds (ms). The presentation of the stimuli was controlled by the software DMDX with a laptop computer.

2.3.2. Non-speech task

After participating in the discrimination of speech tones, the same subjects took part in a discrimination of non-speech 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 of AA pairs was presented ten times. 60 tokens (15 AB pairs × 4 times) of AB pairs and 60 tokens (6 AA pairs ×10 times) of AA pairs were used. The number of AA and AB pairs was balanced, and there were altogether 120 tokens, grouped into 4 blocks.

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.

Some discrepancies between the speech and non-speech tasks, such as the number of AA pair tokens, the speakers who produced the stimuli, and the equipments, can be observed. It is because the speech and non-speech tasks were originally designed for different studies. In order to reduce the effect of these discrepancies, we focus on the AB pairs only in this study, so the different numbers of AA pair tokens in the two tasks should not affect the results seriously. Furthermore, the two tasks used the same paradigm and the same subjects. Therefore, the results of the two tasks can be compared despite some discrepancies.

3. Results

3.1. Speech vs. Non-speech tasks

Both error percentage (EP) and reaction time (RT) were collected. All the participants made few errors for the AA pairs in both speech and non-speech tasks. As mentioned above, the AB pairs will be the focus of the study. Therefore, only the results of the AB pairs are reported here.

The EP and RT (collapsed across presentation order and across the two syllables) of the speech and non-speech tasks are illustrated in Figure 2. The non-speech task had lower EP and shorter RT than the speech task for each L1 group. Among the L1 groups, the Mandarin group performed well with the lowest EP and the shortest RT in both tasks.

Two Repeated-Measures ANOVAs were conducted for EP and RT separately with L1 Group (Mandarin, English, and French) as the between-subject factor and Task (speech task vs. non-speech task) as the within-subject factor. In terms of EP, the main effects of L1 group [$F(2, 29) = 4.53$, $p = 0.019$] and Tasks [$F(1, 29) = 56.34$, $p < 0.001$], and the interaction effect [$F(2, 29) = 3.93$, $p = 0.03$] are significant. Similarly, in terms of RT, the results found the main effects of L1 group [$F(2, 29) = 7.15$, $p = 0.003$], Tasks [$F(1, 29) = 183$, $p < 0.001$], and the interaction effect [$F(2, 29) = 4.25$, $p = 0.024$]. Figure 2 shows that all the L1 groups had a lower EP and a short RT in the non-speech task than in the speech task. Moreover, the Mandarin subjects did the best in terms of EP and RT in the

[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.

speech task, but their performance was similar to the other groups in the non-speech task.

Since the interaction effects of L1 group and task were found for EP and RT, two types of post-hoc tests were conducted to explore the significant effects.

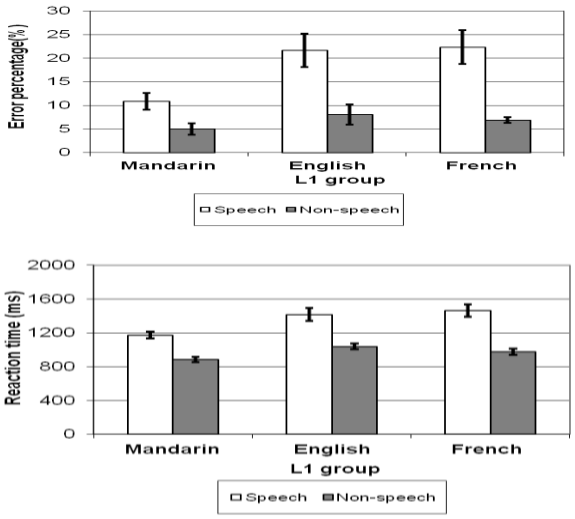

Figure 2. *Error percentage (top panel) and reaction time of correct pairs (bottom panel) in speech and non-speech tasks by different L1 groups.*

First, paired t-tests were conducted for each L1 group to compare their performance in the speech and non-speech tasks in terms of EP and RT. As shown in Table 1, it was found that each L1 group had significantly lower EP and shorter RT in the speech task than the non-speech task.

Table1: *Results of paired t-tests for error percentage and reaction time.*

	EP		
L1 group	Mandarin	English	French
Speech vs. Non-speech	t(11)=5.4 p<0.001	t(11)=8.9 p<0.001	t(11)=8.9 p<0.001
	RT		
L1 group	Mandarin	English	French
Speech vs. Non-speech	t(11)=8.9 p<0.001	t(9)=6.6 p<0.001	t(9)=8.2 p<0.001

Second, four One-Way ANOVA tests with L1 group as a between-subject factor were conducted to investigate whether they had different performance in terms of EP and RT within each task. The Mandarin group did better than the English and French groups in both EP (E: p=0.045; F: p=0.035) and RT (E: p=0.026; F: p=0.007) in the speech task. In the non-speech task, the effect of L1 group was only found in RT, which is mainly due to the difference between the Mandarin and English groups (p=0.007). No significant difference was found between the Mandarin and French groups, and between the English and French groups.

In summary, differences between the speech task and the non-speech task were found for all the L1 groups. While the Mandarin group performed differently from the other groups in the speech task, most of the differences among the three L1 groups in the non-speech task were not statistically significant.

3.2. The effect of individual tone pairs

Since different tonal contrasts were involved in the study, the perceptual performance of each L1 group was examined with respect to individual tone pairs in the speech and non-speech tasks. The data of both RT and EP were analyzed, but only the EP results were reported here due to page limit.

Figure 3 shows the error pattern of each L1 group in the speech and non-speech tasks. The three L1 groups made fewer errors in the non-speech task than those in the speech task for all the tone pairs. Among the L1 groups, the Mandarin group performed well in that they had the lowest EP for most of tone pairs in both tasks.

Two Repeated-Measures ANOVA tests were conducted on the speech and non-speech tasks separately with L1 group as a between-subject factor (3 levels) and tone pairs (15 levels) as a within-subject factor. Regarding the speech task, the results revealed the main effects of L1 group [F (2, 29) =5.1, p=0.012], Tone pairs [F (5.7, 167) =41.3, p<0.001], and the interaction between them [F (11.5, 167) =11.5, p<0.001]. In contrast, the results of the non-speech task only revealed the main effect of Tone pairs [F (3.1, 88.9) =34.9, p<0.001], but the L1 group effect [F (2, 29) =1.13, p=0.34] and the interaction effect [F (6.1, 88.9) = 1.27, p=0.28] were not found in the non-speech task.

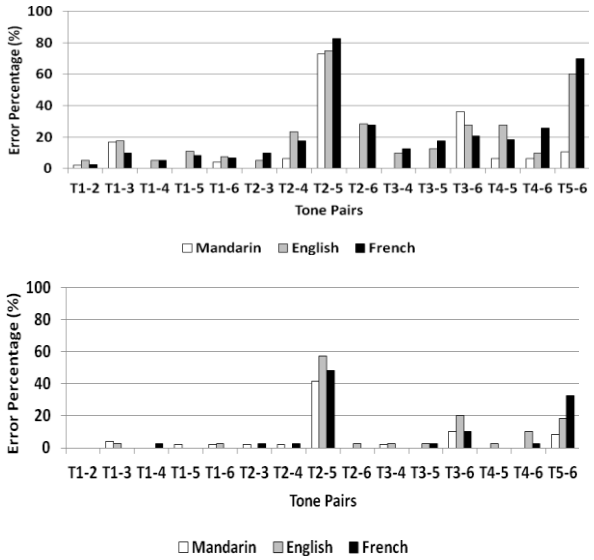

Figure 3. Error percentage (%) of 15 tone pairs by different L1 groups in the speech task (top panel) and non-speech task (bottom panel).

As Figure 3 shows, the speech task and the non-speech task shared some confusable pairs, but differed in specific pairs. On the one hand, T2-T5 was the most difficult pair with the highest EP in both tasks. A similar pattern was also found for the pairs of T3-T6 and T5-T6. The T3-T6 pair had a higher EP than the other level tone pairs, T1-T3 and T1-T6, for all the

L1 groups in the two tasks. The T5-T6 pair appeared to be difficult for the English and French groups in both tasks. Therefore, the difficulty of T2-T5, T3-T6 and T5-T6 in the speech task is attributed to the acoustic similarity of these tones.

On the other hand, the subjects' performance in the speech task differed from that in the non-speech task in terms of specific tone pairs. Although all the L1 groups increased errors from the non-speech task to the speech task, the three L1 groups had difference performance. For example, the Mandarin group had the highest EP in the T3-T6 pair in the speech task, but similar pattern was not found in the non-speech task. The English and French groups had low EP for most tone pairs in the non-speech task, but their error increased a lot for the T2-T6, T4-T5, T4-T6 pairs in the speech task.

To sum up, the effect of L1 group was found by comparing the patterns of individual tone pairs in the speech and non-speech tasks and by testing the different L1 groups' performance in the speech task. The subjects had different patterns of individual tone pairs in the speech and non-speech tasks under the influence of L1expereince.

4. Discussion

In terms of the subjects' overall performance, each L1 group's performance in the speech task was different from that in the non-speech task. The three L1 groups discriminated the non-speech stimuli much better than the speech stimuli, because pure auditory ability was mainly involved in the non-speech task. The non-speech task was much less demanding than the speech task, and all the L1 groups did similarly in the non-speech task. In contrast, a clear different performance between the Mandarin group and the other groups was found in the speech task. The effect of linguistic experience was suggested to differentiate the speech task from the non-speech task.

In terms of the subjects' performance of individual tone pairs in the two tasks, both of the effect of L1 experience and psychoacoustic factors were found.

The three L1 groups shared some confusable tones in the two tasks due to the psychoacoustic similarity of these tones. First, all the subjects found the T2-T5 pair the most confusable because T2 and T5 are acoustically similar and only differ in the magnitude of the final rising movement. Second, T3-T6 was the most difficult to discriminate among the level tones for all the L1 groups. The shorter acoustic distance between T3 and T6 than that between T1 and T3 contributes to the relative difficulty of this pair among the level tone pairs. In addition, T5 and T6 shared the pitch onset, which made this pair confusable in both tasks.

On the other hand, the three L1 groups had different performance in the two tasks under the influence of L1 experience.

Contrary to the English and French groups who made more errors in most of the tone pairs in the speech task, the Mandarin group found most of the tone pairs, including acoustically similar pairs such as the T5-T6 pair, easy to distinguish in all the tasks. The good performance of the Mandarin group can be explained by their linguistic experience with Mandarin tones. However, the Mandarin group found the level tones difficult to distinguish supported by the high EP of the T3-T6 pair in both tasks. As there is only one level tone in the Mandarin tone inventory, the Mandarin speakers lose the sensitivity to the level tone variations, which are within-category differences for them.

The English and French groups did not have significant difference in the two tasks. The results disconfirm the speculation in [7] that French speakers may outperform English speakers because French prosody has " no constraint by lexical accentuation and stress patterns as English does" [7]. Owing to the lack of lexical tones in the native prosodic system, the English and French groups perceived tones in a similar way. Their L1 experience was found to exert an influence by hindering their perception of speech tones, which resulted in much higher EP, especially for contour tone pairs, in the speech task than that in the non-speech task. Additionally, both groups perceived speech tones mainly relying on psychoacoustic aspects of the stimuli, which resulted in the confusion of the T5-T6 pair.

It was suggested in [8] that while Mandarin speakers were sensitive to pitch direction, English and French speakers only placed emphasis on pitch height. The speech task results appeared to agree with it because while the level tones were equally difficult for the Mandarin group to distinguish, the English and French groups found some contour tones confusable in our study. To answer the question of which acoustic cues of the stimuli different L1 listeners would attend to, further study is needed.

5. Conclusion

Both psychoacoustic similarity of stimuli and L1 experience were found to affect the naïve perception of speech tones. While some tones were equally confusable in the speech and non-speech tasks due to their acoustic similarity, L1 experience exerted an influence on the perception of speech tones as differences were found among the L1 groups in terms of overall performance and individual tone pairs.

6. References

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

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

[3] Bent, T., Bradlow, A. R., and Wright, B. A, "The influence of linguistic experience on pitch perception in speech and non-speech sounds", Journal of Experimental Psychology: Human Perception and Performance, 32(1): 97-103, 2006.

[4] Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon,C., Lacerda, F., and Keller, P., "Perception of lexical tone across languages: Evidence for a linguistic mode of processing", in H. T.Bunnell and W. Idsardi [Ed.], Proceedings of the Fourth International Conference on Spoken Language Processing, 1: 2514–2517, 1996.

[5] Zheng, H. Y.,Tsang, P. W-M. and Wang, W. S-Y., "Categorical Perception of Cantonese Tones in Context: a Cross-Linguistic Study.", in Proceedings of Interspeech 2007: 2745-2748, 2007.

[6] Bauer, R. S. and Benedict, P. K., "Modern Cantonese phonology", New York: Mouton de Greyter, 1997.

[7] Halle, P. A., Chang, Y. C. and Best, C. T., "Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners", Journal of Phonetics, 32: 395-421, 2004

[8] Gandour, J. T., "Tone perception in far Eastern languages". Journal of Phonetics, 11: 149-175, 1983.