

PERCEPTION OF CANTONESE TONES BY MANDARIN SPEAKERS

Mengyue Wu¹, Rikke Bundgaard-Nielsen^{2,3}, Brett Baker¹, Catherine T. Best^{2,4}, Janet Fletcher¹

¹ The University of Melbourne, ² MARCS Institute, University of Western Sydney, ³ La Trobe University,
⁴ Haskins Laboratories

mengyuew@student.unimelb.edu.au; r.bundgaardnielsen@uws.edu.au; bjbaker@unimelb.edu.au; c.best@uws.edu.au;
j.fletcher@unimelb.edu.au

ABSTRACT

The current study investigates how non-native tones are perceived by speakers whose own native language has fewer tones. The analysis is presented within a Perceptual Assimilation Model—Supra-segmental (PAM-S) framework [16]. The results are consistent with PAM-S predictions and indicate both phonetic and phonological assimilation of Cantonese tones by Mandarin speakers.

Keywords: Non-native speech perception; Lexical tones; Perceptual Assimilation Model.

1. INTRODUCTION

That the perception of non-native tones is influenced by native tone language experience is well established (cf. [2, 7, 12, 14, 16, 17]). However, whether this native experience facilitates or interferes with non-native perception remains unclear. Research suggests this depends on both discrepancies and similarities between the L1 and L2 tone systems. How native experience with a comparatively simple tone system might influence listeners' perceptions of more complex non-native tones is also unclear [9].

The Perceptual Assimilation Model (PAM) [1] has been increasingly extended to the supra-segmental level (PAM-S [16]), in particular to tone studies (cf. [4, 6, 8, 10, 11, 12, 15]). PAM-S makes clear predictions about the discriminability of *non-native* tones on the basis of their categorisation (or lack thereof) into the available *native* tone categories. If the non-native tone can be categorised into the native system, it is *categorised*; if not, it is *uncategorised*. When two tones in a non-native contrast pair are perceived as belonging equally well/poorly to a single native category ('Single Category', SC), discrimination is predicted to be poor. When two non-native tones are categorised into the same native category, but with one as a better instance than the other ('Category Goodness', CG), discrimination is expected to be good. Finally, when two tones belong to two separate native categories ('Two-Category', TC), discrimination is expected to be excellent. PAM-S also provides predictions for contrasts involving uncategorised tones: 'Uncategorised-Uncategorised'

(UU) and 'Uncategorised-Categorised' (UC) contrasts. UU and UC can be further categorised into different sub-types differing in discriminability. When both tones are categorised into the same set of native categories, the contrast is 'same set' (UC-s). When there is a partial overlap in the perceived similarity to native categories, it is 'partial overlap' (UC-o). With no perceived overlap, it is 'no overlap' (UC-no). UC-no contrasts are predicted to be easy to discriminate, UC-o contrasts are more difficult, and UC-s are the most difficult.

Previous studies have found that L2 tones are mapped onto tone language speakers' native tone categories, and that this native influence is difficult to overcome, even with training [8]. Research also suggests that some universal difficulties in discrimination might be due to psychoacoustics [2, 13]. Support for PAM predictions exists regarding the discriminability of TC and CG contrasts [9, 10, 15]. However, different categorisation methods greatly influence individual study results.

Typical segmental perception studies differentiate between phonetic and phonological assimilation. Phonetic assimilation occurs when one L2 phone is perceived as the phonetic equivalent of the L1 category. In contrast, phonological assimilation occurs when the same phonological behaviour is evident in both the L1 and L2 categories, i.e. an L2 tone is perceived as the allophonic variant of an L1 tone. Few studies have examined this issue in terms of supra-segmental information. A recent study [18] has suggested that phonological assimilation only occurs in experienced listeners, while other findings indicate that phonological assimilation may also occur in inexperienced listeners [11, 14].

2. THE PRESENT STUDY

The present study investigates Cantonese tone perception by Mandarin speakers. Cantonese has six tones. Three of these are level tones: high-level 55; mid-level 33 and low-level 22 in the Chao system [3]. Two are rising tones: a high-rising 25 and a low-rising 23. The last is a low-falling, 21. In contrast, Mandarin has four tones. These are a high-level tone 55, a high-rising tone 35, a falling-rising (dipping) tone 214 and a high-falling tone 51. Unlike Cantonese, the tone

contrasts in Mandarin are differentiated primarily by pitch trajectory, rather than pitch height.

The main objectives of the current study are to examine how speakers of a language with a relatively simpler tone system perceive tones from a more complex system, and whether native experience with native tones facilitates or hinders non-native tone perception. The study also aims to determine whether both phonetic and phonological assimilation occur in inexperienced listeners. The experiments consist of a categorisation task plus goodness rating (GR) and a discrimination task.

2.1. Experiment 1: Tone Categorisation

This experiment tests how Mandarin speakers categorise Cantonese tones to their native system.

2.1.1. Method

Participants Twenty native Beijing Mandarin speakers (mean (M) age 23.8 years, standard deviation (SD) = 2.85) participated. They had neither proficiency in Cantonese, i.e., were inexperienced, nor any extensive musical training. All participants passed a pure-tone hearing screening (250–8000 Hz at 25dB HL).

Stimuli A total of 18 tokens (three repetitions of /mɔː/ using all six Cantonese tones) were recorded by a female native Cantonese speaker (25.6 years old). Stimulus recording was done at the MARCS Auditory recording booth, with a sampling rate of 44010Hz, and a resolution of 16 bits. We used /mɔː/, as it forms real words in combination with all four Mandarin tones and resembles the real Cantonese word ‘mo’ /mɔː25/, ‘touch’, which has a high rising tone.

Procedure Participants were asked to categorise randomised presentations of 120 trials of the target /mɔː/ (6 tones × 20 repetitions) into the four Mandarin tone categories: level, rising, dipping and falling

tones (presented as 摸, 魔, 抹, 末). Additionally, participants were given an ‘unknown’ option. Listeners were instructed to compare the target syllable with the four Mandarin words, choosing the most similar and inputting a Goodness Rating (GR: 1 [poor] to 5 [excellent]). They were instructed to choose ‘unknown’ when they could not identify a target word’s tone with any of the native tone categories provided.

2.1.2. Results

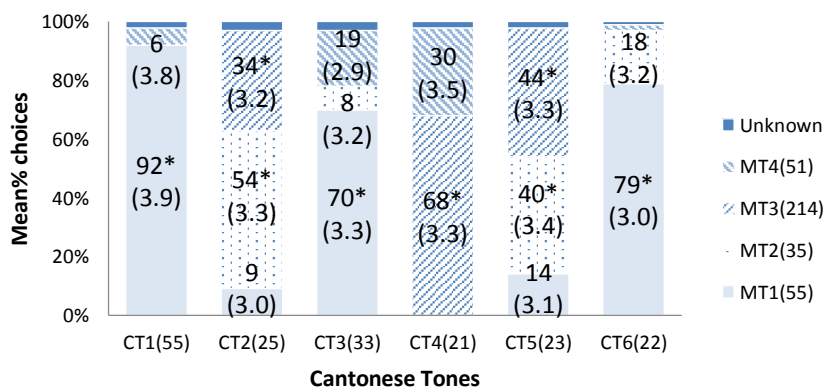
The categorisation results are presented in Figure 1. This paper adopts the ‘categorisation’ criteria proposed by [16]. Here, an L2 tone has to satisfy two criteria to count as ‘categorised’: the frequency of the selected category must be significantly higher than both chance level and other category selections. In the present study, the chance level for each tone is 20%.

As Figure 1 shows, the only level tone available in Mandarin (MT = Mandarin tone; 55 high-level) was the primary choice for all three level Cantonese tones (CT = Cantonese tone). In 92% of cases, CT1 (55) was categorised as the Mandarin high-level tone (GR = 3.9). The Mandarin level tone was chosen 70% of the time for CT3 (33; GR 3.3) and 79% of the time for CT6 (22; GR 3.0). The two Cantonese rising tones, CT2 (25) and CT5 (23), were categorised into MT2 (35) or MT3 (214). However, MT2 (35) was the primary choice (54%) for CT2 (25), chosen significantly more often than MT3 (214; 34%). In contrast, the same two tones MT2 (35; 40%) and MT3 (214; 44%) were selected with almost equal frequency for CT5 (23). Finally, CT4 (21) was categorised as MT3 (214; 68%) of the time (GR 3.3). Interestingly, MT4 (51) was selected 30% of the time, with a slightly higher GR of 3.5.

A chi-square test revealed a significant association between Cantonese tones and the chosen Mandarin categories ($\chi^2(20) = 2425.146, p < .001$).

Figure 1: Mandarin listeners’ tonal categorisation percentage for each Cantonese tone, with GR in brackets.

Asterisked percentages ($p < .001$) indicate a mean significantly above the chance level (20%).



This was further examined in a two-way repeated-measures ANOVA (CT × MT), which found a significant main effect of CT, $F(5,14) = 45.178, p <$

.001. Additionally, it found a significant effect of MT, $F(3, 285) = 106.065, p < .001$. The CT × MT

interaction was also significant, $F(6, 285) = 246.359$, $p < .001$.

According to the categorisation criteria, all Cantonese tones, with the exception of CT5 (23), were categorised. For CT1 (55), CT3 (33) and CT6 (22), which were all mapped onto the same category, three t-tests were performed on the GRs. As the GRs all differed significantly ($p < .001$), the tone pairs formed by CT1, CT3 and CT6 were considered CG contrasts, rather than SC. Tone contrasts formed with CT5 were considered UC, of either the (s), (o) or (no) type, depending on the identity of the other tone in the contrast pair.

2.1.3. Discussion

The results indicate that in most cases, non-native tones are categorised as the most acoustically similar native counterparts. That all three level tones are categorised as the only level tone available in Mandarin demonstrates that even partial similarity can stimulate phonetic assimilation. However, the differences in GRs suggest that Mandarin speakers are indeed able to differentiate F0 height: Mandarin listeners found CT1 (55) the best fit for MT1 (55), although they also chose MT1 for CT3 (33) and CT6 (22). With the two rising tones—CT2 (25) and CT5 (23), listeners chose MT2, a rising tone (35). However, they also chose the falling-rising MT3 (214), which has an allophonic rising form with the tone pattern (14).

When the Cantonese rising tone (25) is categorised as the rising tone (35) in Mandarin, this again suggests that assimilation occurs at the phonetic level. However, when the rising tone is assimilated to the Mandarin falling-rising (214) tone, this indicates phonological assimilation. Mandarin listeners apparently apply their native phonological knowledge in the following way. They perceive that the falling and rising tones are allophonic variants of the falling-rising tone, in order to categorise the Cantonese rising tone (23) and the Cantonese falling tone (21). This result aligns with [13], but contradicts [18], where no phonological assimilation occurred.

2.2. Experiment 2: Tone Discrimination

The current experiment examines native and non-native discrimination of Cantonese tones.

2.2.1. Method

Participants The same participant group from Experiment 1 participated in this task. In addition, 20 native Hong Kong Cantonese speakers (M age=23.9 years old) was recruited to provide a baseline for the discrimination scores.

Stimuli The stimuli used in the discrimination task were three carrier syllables: /ba:p/, /bi:/ and /bu:/. These were selected as none of the six tones in these syllables form a real Cantonese word. A total of 108 stimuli (3 syllables \times 3 tokens per tone \times 6 tones \times 2 repetitions) were recorded by two female speakers, under recording conditions identical to those of Experiment 1. The stimuli were screened by three native Cantonese speakers, who verified these tones were produced correctly.

Procedure An AXB discrimination task was presented using E-Prime 2.0. During the task, participants had to decide whether the second tone presented was the same as the first tone, or the last tone. The task consisted of 360 trials (3 syllables \times 2 speakers \times 15 contrasts \times 4 triad types [AAB, ABB, BAA, BBA]), divided into three blocks by syllable. Within each trial, the same speaker was used.

2.2.2. Results

The discrimination results are summarised in Figures 2 and 3. The Cantonese listeners' M score (93%) was significantly higher than that of the Mandarin listeners (78%).

Figure 2: Mean correct discrimination (in %) for each Cantonese tone pair by Mandarin listeners.

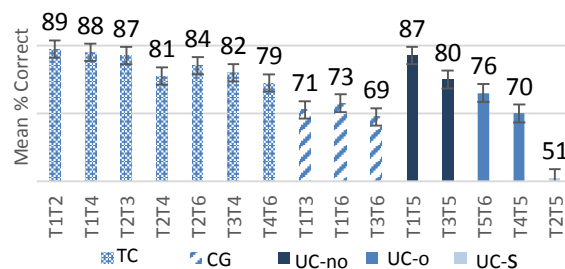
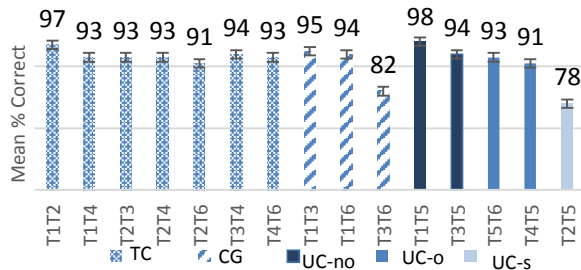


Figure 3: Mean correct discrimination (in %) for each Cantonese tone pair by Cantonese listeners.



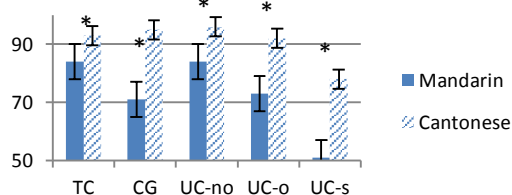
A mixed-factors ANOVA—with group as the between-subjects factor and tone pair as the within-subjects factor—showed a significant effect of group, $F(1, 36) = 2359.507$, $p < .001$, as well as tone pair, $F(14, 504) = 228.988$, $p < .001$, and the Group \times Tone Pair interaction, $F(14, 504) = 70.143$, $p < .001$. The ANOVA for the Mandarin group showed a main

effect of tone pair, $F(14, 270) = 174.686, p < .001$. Post hoc t-tests, with Bonferroni corrections for multiple comparisons, further revealed that scores for the contrasts classified as CG—T1-T3 (71%), T1-T6 (73%) and T3-T6 (69%)—were significantly lower ($p < .001$) than those for TC contrasts. UC-(no) pairs—T1-T5 (87%) and T3-T5 (80%)—were not significantly lower than the TC pair. While one of the UC-overlap pairs (T4-T5) was significantly lower than the TC contrasts, the other was not. The UC-s pair T2-T5 (51%) was significantly lower than all other tone pairs, $p < .001$.

The discrimination for TC assimilated tone contrasts was 84% for Mandarin and 93% for Cantonese groups (see Fig. 4). The CG contrast difference is larger (71% for the Mandarin and 95% for Cantonese groups). Within the UC contrasts, the groups also showed different patterns on the three sub-sets. For UC-no, Mandarin speakers performed at 84%, while native discrimination was 96%; for UC-overlap, the discrimination were lower (73%) for Mandarin speakers, but still quite high (92%) for Cantonese speakers.

The most difficult case for both groups was the UC-s contrast, where Mandarin speakers did not exceed chance level (51%). Cantonese speakers achieved just 78% for this contrast. The results confirm this study’s predictions: that TC contrasts should allow excellent discrimination, while CG contrasts should display moderate to good discrimination. Further, within UC groups, a UC-no contrast should allow better discrimination than a UC-o contrast. The UC-s contrast, which involves activating the same set of native categories, is the most difficult to discern for the non-native group.

Figure 4: Mean discrimination of the contrast types. The asterisk (*) indicates the difference between the two listener groups is significant ($p < .001$).



2.2.3. Discussion

The present study shows that Mandarin speakers discriminate Cantonese tones moderately well. This is probably due to their native language experience with tone. However, their difficulties with some particular non-native tone pairs suggest that the native system also hinders non-native tone perception. These results support PAM-S predictions and a number of previous findings: TC contrasts are better

discriminated than CG contrasts, while the discriminability of UC contrasts varies greatly. The latter depends on how the contrasts are perceived as overlapping (or not) with native categories [14]. The results also highlight that the selected categorisation criterion of a given study systematically affects predictions about perception. For example, [13] proposed that the Mandarin tone pair T2-T3 should be a CG contrast for Cantonese listeners, while [6] classified this pair as a TC contrast, even though poor discrimination occurred in both cases. In the current study, the Cantonese tone pair T2-T5 was classified as ‘UC-s’, leading to a prediction of poor discrimination, supported by our findings. This contradicts [9], where the Cantonese T2-T5 contrast was classified as CG, resulting in a prediction of moderate to very good discrimination.

For both native and non-native listeners, the results indicate that discriminating the two rising tones (T2-T5) is the most difficult task. This accords with previous studies indicating that native adult Cantonese speakers sometimes find this distinction difficult, as do speakers of French and English [6, 13]. This suggests that the T2-T5 contrast may be more difficult to perceive than other contrasts, regardless of the listener’s language background. The second most difficult contrast for both groups is T3-T6, which can be explained in terms of the acoustic properties of the tones: T3 (33) and T6 (22) have the smallest F0 difference, and no difference in trajectory. The other two easily confused tone pairs are T4 (21)-T5 (23) and T4 (21)-T6 (22), as they share similar F0 onsets with only slightly different F0 offsets.

3. CONCLUSION

The current study provides a number of new insights into cross-language tone perception in situations where native speakers of relatively simple tonal systems are presented with a more complex system. Primarily, our results indicate that native tone experience can be a double-edged sword when applied to cross-language tone perception. Native experience both helps and hinders the perception of a novel tone system.

The results also expand on previous research in two ways. Firstly, the results demonstrate that phonological (in addition to phonetic) tone processing is present in inexperienced listeners. This is an important insight, with considerable implications for phonetic and phonological second language tone learning. Secondly, in relation to methodology, the results show that even small differences in the categorisation criteria chosen for a given study will have significant implications for interpreting results.

4. REFERENCES

- [1] Best, C. T. 1995. A direct realist perspective on cross-language speech perception. *Speech Perception and Linguistic Experience: Theoretical and methodological issues in cross-language speech research*. York: Timonium, 171–204.
- [2] Burnham, D., Kasisopa, B., Reid, A., Luksaneeyanawin, S., Lacerda, F., Attina, V., Webster, D. 2014. Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholinguistics*, 1–33.
- [3] Chao, Y. R. 1930. A system of tone letters. *Le Maître Phonétique*, 45, 24–27.
- [4] Chiao, W.-H., Kabak, B., Braun, B. 2011. *When more is less: Non-native perception of level tone contrasts* <http://ling.unikonstanz.de/pages/home/braun/articles/Chiao.Kabak.Braun-1.pdf>
- [5] Guion, S. G., Pederson, E. 2007. Investigating the role of attention in phonetic learning. *Language experience in second language speech learning*, 57–77.
- [6] Hao, Y.-C. 2011. Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *J of Phonetics*, 40(2), 269–279.
- [7] Lee, Y.-S., Vakoch, D. A., Wurm, L. H. 1996. Tone perception in Cantonese and Mandarin: a cross-linguistic comparison. *J of Psycholinguistic Research*, 25, 527–542.
- [8] Leung, A. S. 2008. Tonal assimilation patterns of Cantonese L2 speakers of Mandarin in the perception and production of Mandarin tones. *Proc. 2008 CLA*.
- [9] Qin, Z., Mok, P. P. 2011. Discrimination of Cantonese tones by speakers of tone and non-tone languages. *Kansas Working Papers in Linguistics*, 34.
- [10] Reid, A., Burnham, D., Kasisopa, B., Reilly, R., Attina, V., Rattanasone, N. X., Best, C. T. 2014. Perceptual assimilation of lexical tone: the roles of language experience and visual information. *Attention, Perception & Psychophysics*, 1–21.
- [11] So, C. K. 2012. Cross-language categorization of monosyllabic foreign tones: effects of phonological and phonetic properties of native language. In: Stolz, T., Nau, N., Stroh, C. (eds), *Monosyllables: From phonology to typology*. Berlin, Germany: Akademie Verlag, 55–69.
- [12] So, C. K., Best, C. T. 2008. Do English speakers assimilate Mandarin tones to English prosodic categories? *Proc. 8th Interspeech, incorporating the 12th SST*. Baixas, France.
- [13] So, C. K., Best, C. T. 2010 a. Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, 53, 273–293.
- [14] So, C. K., Best, C. T. 2010 b. Discrimination and categorization of Mandarin tones by Cantonese speakers: The role of native phonological and phonetic properties. *Proc. 13th SST*. Melbourne, Australia.
- [15] So, C. K., Best, C. T. 2011. Categorizing Mandarin tones into listeners' native prosodic categories: The role of phonetic properties. *Poznań Studies in Contemporary Linguistics*, 47, 133–145.
- [16] So, C. K., Best, C. T. 2014. Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, 36 (2), 195–221.
- [17] Wayland, R. P., Guion, S. 2004. Training English and Chinese listeners to perceive Thai tones: a preliminary report. *Language Learning*, 54, 681–712.
- [18] Wu, X., Munro, M. J., Wang, Y. 2014. Tone assimilation by Mandarin and Thai listeners with and without L2 experience. *J of Phonetics*, 46, 86–100.