Effects of linguistic experience on the perception of high-variability non-native tones

Yung-hsiang Shawn Chang, Yao Yao and Becky H. Huang

Citation: The Journal of the Acoustical Society of America **141**, EL120 (2017); doi: 10.1121/1.4976037
View online: https://doi.org/10.1121/1.4976037
View Table of Contents: https://asa.scitation.org/toc/jas/141/2
Published by the Acoustical Society of America

Effects of linguistic experience on the perception of high-variability non-native tones

Yung-hsiang Shawn Chang[a)]
Department of English, National Taipei University of Technology, 1, Sec. 3, Zhongxiao East Road, Taipei, Taiwan
shawnchang@ntut.edu.tw

Yao Yao
Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
ctyaoyao@polyu.edu.hk

Becky H. Huang
Department of Bicultural-Bilingual Studies, University of Texas at San Antonio, One UTSA Circle, San Antonio, Texas 78249, USA
becky.huang@utsa.edu

Abstract: Whether tone language experience facilitates non-native tone perception is an area of research that previously yielded conflicting results, potentially because of the lack of systematical control of speaker normalization effects across studies. Under a high-variability testing condition with controlled speaker normalization cues, Cantonese (native controls), Mandarin (Cantonese-naive tone listeners), and English (non-tone listeners) listeners identified three Cantonese level tones. The results indicate a facilitatory effect of tone experience on non-native tone perception when normalization for inter-speaker variation is required.

[RS]
Date Received: September 20, 2016 **Date Accepted:** January 24, 2017

1. Introduction

Previous cross-language research (e.g., Gandour and Harshman, 1978; Gandour, 1983; Hallé *et al.*, 2004) has established the importance of linguistic experience in the perception of lexical tones: native tone language speakers process $F0$ differently from non-tone language speakers, which accounts for non-tone speakers' poorer performance in tone perception in general. In addition to comparing tone perception by speakers of a tone language and a non-tone language, researchers have further explored whether speakers of a tone language have an advantage over non-tone language speakers in perceiving tonal contrasts in a different tone language, but the findings were mixed. While some studies reported that tone speakers outperformed non-tone speakers in non-native tone perception (Liang and van Heuven, 2007; Wayland and Guion, 2004), others found no significant performance difference between the two groups (Francis *et al.*, 2008; Hao, 2012; So and Best, 2010). Yet other researchers have found selective advantages of tone language experience in the perception of non-native tones (Lee *et al.*, 1996; Qin and Jongman, 2015; Qin and Mok, 2011).

These conflicting findings have been discussed in terms of differing complexities of tone inventories. Lee *et al.* (1996) found that while their Cantonese listeners were better than English listeners at discriminating Mandarin tones, Mandarin listeners did not perform better than English listeners in Cantonese tone discrimination. They claimed that the asymmetric tone experience effects were due to the fact that Cantonese has more tones (six) than Mandarin (four) and is therefore harder to acquire and perceive than Mandarin. The other factor that has been addressed is tonal type, especially concerning the difference between level tones—which contrast in pitch height—and contour tones—which contrast in both pitch height and pitch direction. Qin and Mok (2011) and Qin and Jongman (2015) found that Mandarin listeners had better performance than English listeners in distinguishing Cantonese contour-level and contour-contour tone pairs; conversely, English listeners performed better than Mandarin listeners in distinguishing Cantonese level-level tone pairs. To account for

[a)]Author to whom correspondence should be addressed.

these results, Qin and Mok (2011) and Qin and Jongman (2015) argued that tone listeners (i.e., Mandarin listeners) are more sensitive to *F*0 direction, whereas non-tone listeners give more perceptual weighting to *F*0 height.

Another factor that may have rendered conflicting cross-study results, but has received little attention, is the amount of speaker normalization involved in the perceptual task. In a tone recognition task, speaker normalization is required when the stimuli presented are from multiple speakers (So and Best, 2010; Wayland and Guion, 2004) but not when all the stimuli come from a single speaker (used in most of the aforementioned studies); furthermore, when speaker normalization is indeed required, it may be facilitated by presenting the tone stimuli in context (Francis *et al.*, 2008; Hao, 2012; Liang and van Heuven, 2007) as opposed to presenting the tones in isolation (as in most of the aforementioned studies). To our knowledge, Lee *et al.* (2009) is the only available study that systematically compares native tone listeners' and non-tone listeners' perception of tones produced by single vs multiple speakers and tones presented in isolation vs in context. They found that while speaker variability compromises native and non-tone listeners' performance in a similar manner, contextual information benefits native tone listeners more than non-tone listeners. Nevertheless, the impact of speaker variability and contextual information on tone listeners' perception of non-native tones remains underspecified. In this regard, to control for speaker normalization effects, the current study adopted a high-variability testing paradigm from Lee *et al.* (2011, 2014). In this paradigm, the auditory stimuli were produced by a large number of speakers (so that listeners could not be familiar with a particular voice) and presented in isolation (so that there was no syllable-external *F*0 cues for speaker normalization); as such, tone perception has to solely rely on syllable-internal *F*0 cues and inter-speaker tone variation has to be normalized by estimating speakers' individual *F*0 ranges.

We specifically examined the perception of the three level tones in Cantonese by Mandarin and English listeners because (1) the only perceptual cue to Cantonese level tone perception is *F*0 height. Without having to address the interaction of multiple *F*0 cues that characterize contour tones, the cross-linguistic perception results can be more easily interpreted, and tone language experience effects more clearly observed, and (2) it is a topic where conflicting effects of language background [Qin and Mok (2011) and Qin and Jongman (2015) vs Francis *et al.* (2008)[1]] were reported.

Using a high-variability testing paradigm we assessed how tone language experience modulates listeners' ability to normalize for speaker variability when acoustic cues typically used for speaker normalization (e.g., contextual information, repeated exposure to the same voice) are absent. If Cantonese-naive Mandarin listeners perform better than English listeners, it will indicate that experience with lexical tones and normalizing for inter-speaker tone variation can facilitate the perception of high-variability non-native tones. If Mandarin listeners perform worse than English listeners, it will suggest that Mandarin listeners have disregarded level-tone variations (i.e., less sensitive to pitch height differences) and assimilated the three Cantonese tones into their native level tone (Mandarin Tone1).

2. Method

Two sets of Cantonese words were selected for the current experiment: /se1/ (high-level tone; "some"), /se3/ (mid-level tone; "spill"), /se6/ (low-level tone; "shoot"), as well as /fu1/ ("husband"), /fu3/ ("rich"), and /fu6/ ("negative"). Thirty-four native speakers of Hong Kong Cantonese, recruited from the Hong Kong Polytechnic University, produced each of these words embedded in a carrier phrase *ngo5 duk6* __ *zi6* "I read the word___" three times in random order. All materials were presented in traditional Chinese characters.

In view of a trend of a merger between the mid- and low-level tones in Hong Kong Cantonese reported in Mok *et al.* (2013), the *F*0 contours of the stimuli were plotted by speaker and visually inspected. The speakers whose productions indicated overlapping mid- and low-tones were excluded. After screening, 24 speakers (12 males, 12 females) were found to exhibit a clear three-way contrast in Cantonese level tones. From these 24 speakers' productions, the last rendition of each word was used as the stimuli, creating a total of 144 stimuli. Peak amplitude of all stimuli was normalized to 65 dB. Figure 1 displays the *F*0 traces of the stimuli produced by male and female speakers, respectively. Substantial tonal overlap is evident within both gender groups. Table 1 shows the mean *F*0 values and duration of each word produced by male vs female speakers. It can be seen that female speakers have larger *F*0 ranges (62 Hz) than males (35 Hz). Female speakers' level tones are also more distinct (42 Hz difference between high- and mid-tones; 20 Hz difference between mid- and low-tones) than those in male productions (23 Hz; 12 Hz). In terms of syllable duration, high tones appeared to be shorter than the other two tones. We conducted a series of paired

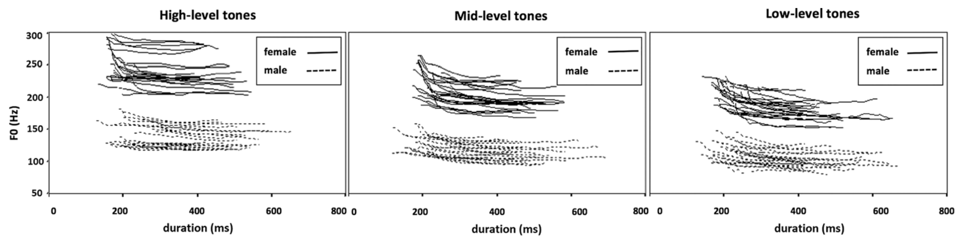

Fig. 1. $F0$ traces of the level tone stimuli (collapsed across syllables—/se/ and /fu/).

t-tests to evaluate whether duration differences among the three tones were significant. The results showed that only the high-level /fu/ stimuli produced by females were marginally significantly shorter than the low-level counterparts ($p = 0.052$). Therefore, syllable duration should not be a useful cue for level tone identification.

Thirty Cantonese listeners (native controls), 30 Mandarin listeners (tone speakers that are naive to Cantonese, as self-reported on the language background form), and 30 English listeners (non-tone speakers), all aged between 18 and 30, were, respectively, recruited and tested at the authors' affiliated institutions. The participant selection criteria were: none of the Mandarin listeners had learned Cantonese; none of the English listeners had learned a tone language before. In addition, all participants self-reported normal speech and hearing, and no participants had received formal musical training.

To ensure that identification of pitch height of Cantonese level tones was not confounded by pitch perception deficits, all participants first took an adaptive pitch test at the Tonometric website (tonometric.com). The test determined the smallest pitch difference (in Hz) one could consistently discriminate. We used 12 Hz, the smallest difference between the average $F0$ values of the three tones (see Table 1), as a screening threshold for all participants.

Prior to the tone identification task, the participants took a 5-min familiarization session on the computer. They were explained the function of lexical tones, and the contrastive pitch heights of three Cantonese level tones. They also saw the pitch tracks of six speech samples of the /si/ words (3 level tones * 2 speakers—one male and one female) and listened to these samples. The participants did not perform any identification task in this session; they only passively viewed the explanation and listened to the speech samples.

The participants then proceeded to a three-way forced-choice identification task in E-Prime (v2.0; Psychological Software Tools, Pittsburgh, PA). The 144 stimuli were assigned to 6 blocks (3 blocks of /se/ and 3 blocks of /fu/), and each block contained 24 stimuli—one from each speaker—that were balanced in tone height. A randomized presentation order (of the six blocks and of the stimuli within each block) was used for each participant. Each stimulus was played once. On the instruction page, the participants were told that they would be listening to six Cantonese words. Their task was to identify the sound by choosing one of the three tone labels "High," "Mid," or "Low" (for non-Cantonese listeners) or three characters (for Cantonese listeners). The participants were told that their responses were timed and that they should respond as quickly as possible. Participants were given six practice trials with the /se/ and /fu/ words recorded by six speakers not used in the actual test. In addition to

Table 1. Mean (with SD) $F0$'s (in Hz) and duration (in ms) of Cantonese level tones across speaker gender and syllable type.

Gender	Tone	Syllable	$F0$	Duration
Male	High	/se/	136 (14)	427 (91)
		/fu/	139 (17)	414 (59)
	Mid	/se/	114 (12)	449 (68)
		/fu/	116 (13)	472 (92)
	Low	/se/	102 (11)	470 (76)
		/fu/	104 (12)	467 (130)
Female	High	/se/	242 (28)	414 (59)
		/fu/	244 (30)	405 (46)
	Mid	/se/	200 (17)	452 (76)
		/fu/	202 (21)	443 (54)
	Low	/se/	181 (16)	437 (59)
		/fu/	181 (16)	463 (94)

identification decision, reaction time was also measured to further investigate tone processing in relation to linguistic experience and speaker variability (in light of Lee *et al.*, 2009). A possible accuracy-speed tradeoff (e.g., perceptual differences shown not in accuracy, but in response time) can be discussed with inclusion of the reaction time measure. Since reaction time may be confounded by durational differences across tone stimuli, we measured it from the offset of the stimuli to the time a response is made. The response and reaction time data were logged automatically in E-Prime. All the aforementioned tasks were conducted in a quiet room under the experimenter's supervision.

3. Results

The results of the adaptive pitch test showed that all 90 participants were able to discriminate tones that were 12 Hz or less apart. In other words, pitch height differences among our Cantonese level tone stimuli should be perceivable to all participants. Identification results showed an overall accuracy [with standard deviation (SD)] of 55% (13.6%) for Cantonese listeners, 52.9% (20.2%) for Mandarin listeners, and 43.9% (18.4%) for English listeners. Native controls' low performance suggests that great overlap of tone categories, as a result of inter-speaker variations (see Fig. 1), renders tone identification challenging even for native listeners.[2] The mean reaction time (in ms with SD) was 1152 (523) for Cantonese listeners, 1069 (466) for Mandarin listeners, and 1143 (476) for English listeners. As shown by the results, the native controls had slightly better identification performance than Mandarin listeners, who in turn were better than English listeners, although the same ranking was not observed in reaction time. Figure 2 shows the mean accuracy rates and reaction times of each group of listeners, separated by speaker gender and tone height.

To confirm whether listener groups, tone height, and speaker gender affected identification accuracy, accuracy data were analyzed with a mixed-effects logistic regression model using the lme4 package in R. Mixed-effects modeling was chosen because it allows "participant" to be added as a random factor, such that by-subject variability in the effects of fixed factors (i.e., listener group, tone height, and speaker gender) on identification performance can be modeled. Additionally, block (to address whether performance varied across blocks since the degree of tonal overlap was not controlled across the six blocks) and order of blocks (to investigate whether there was a practice effect) were also added as fixed factors. The analysis started with an intercept-only model and then added the fixed factors one by one. The factors and factor interactions were retained if the model was significantly improved. The best-fitting model (see Table 2) included group [*Wald* $\chi^2(2) = 44.171$, $p < 0.001$], tone height [*Wald* $\chi^2(2) = 13.953$, $p < 0.001$], speaker gender [*Wald* $\chi^2(1) = 18.262$, $p < 0.001$], as well as interaction effects of group*tone height [*Wald* $\chi^2(4) = 15.021$, $p < 0.01$], tone height*speaker gender [*Wald* $\chi^2(2) = 396.17$, $p < 0.001$], and group*tone height*speaker gender [*Wald* $\chi^2(6) = 81.674$, $p < 0.001$].[3] Pairwise comparisons revealed that (1) Cantonese listeners were better at identifying high-tones produced by males and low-tones produced

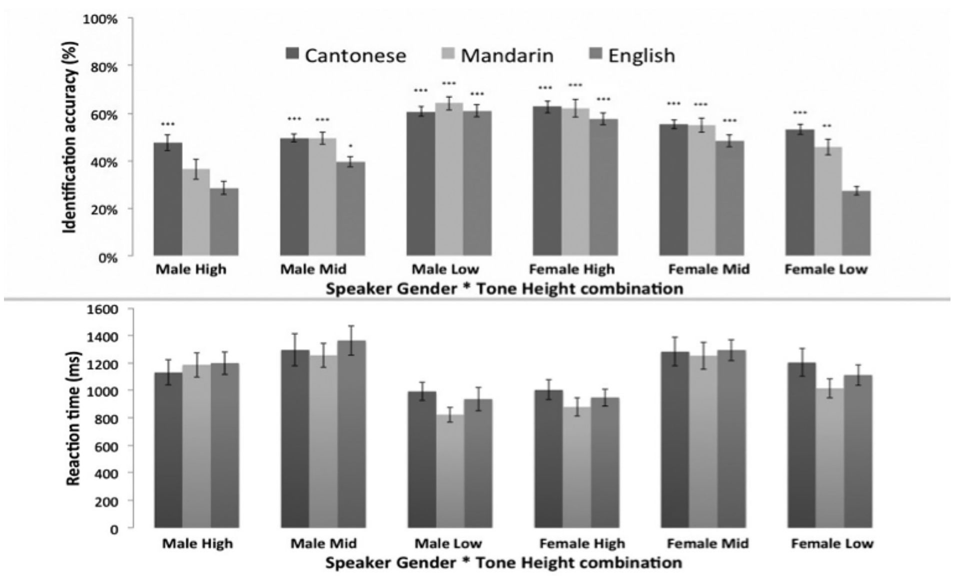

Fig. 2. Mean accuracy (upper panel) and reaction time (lower panel) of identification responses (with standard errors) separated by speaker gender and tone height. Asterisks indicate that identification accuracy was significantly above the chance level of 33.3% (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

Table 2. Summary of the best-fitting mixed-effects logistic model for identification accuracy (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

Effects	Measures		
	Estimated coefficient	Std. error	*z*-value
(intercept)	0.522	0.085	6.162***
Group (TW)	−0.119	0.065	−0.999
Group (US)	−0.280	0.118	−2.362*
Tone (L)	−0.313	0.108	−2.903**
Tone (M)	−0.302	0.108	−2.798**
Gender (M)	−0.612	0.108	−5.678***
Group (TW):Tone (L)	−0.258	0.152	−1.696
Group (US):Tone (L)	−0.906	0.156	−5.806***
Group (TW):Tone (M)	0.267	0.153	1.751
Group (US):Tone (M)	0.088	0.151	0.581
Tone (L):Gender (M)	0.783	0.152	5.158***
Tone (M):Gender (M)	0.381	0.151	2.517*
Group (TW):Tone (H):Gender (M)	−0.348	0.153	−2.273*
Group (US):Tone (H):Gender (M)	−0.552	0.155	−3.555***
Group (TW):Tone (L):Gender (M)	0.683	0.153	4.463***
Group (US):Tone (L):Gender (M)	1.261	0.156	8.075***
Group (TW):Tone (M):Gender (M)	−0.154	0.151	−1.023
Group (US):Tone (M):Gender (M)	−0.224	0.151	−1.483

by females than both Mandarin (both Tukey-adjusted $p < 0.01$) and English listeners (both Tukey-adjusted $p < 0.001$). (2) Mandarin listeners performed better than English listeners in the identification of high- and mid-tones produced by males (both Tukey-adjusted $p < 0.05$) as well as mid- and low-tones produced by females (Tukey-adjusted $p < 0.05$ and $p < 0.001$, respectively). (3) The three groups had comparable performance in the identification of low-tones produced by males and high-tones produced by females. (4) Across all three groups, identifications of low-tones produced by males and high-tones produced by females, though not significantly different from each other, were both significantly better than all other tone height*speaker gender combinations.

We also evaluated whether identification accuracy was above chance (33.3%) by conducting 18 (3 listener groups * 3 level tones * 2 speaker genders) one-sample *t*-tests (Bonferroni correction applied). The results (levels of significance indicated in Fig. 2) showed that only Cantonese listeners were able to identify all three tones produced by both males and females above chance. Mandarin listeners identified high-tones produced by males at chance level, while English listeners identified high-tones produced by males and low-tones produced by females at chance level.

In the reaction time analysis, only the data of correctly identified stimuli were included. Reaction time was logged-transformed prior to analysis with a linear mixed-effects model, with the same fixed and random effects introduced in the accuracy analysis. The best-fitting model presented in Table 3 revealed reaction time was significantly influenced by tone height and the tone height*speaker gender interaction. The *post hoc* analysis exploring the interaction effect indicated that (1) in the identification of tones produced by males, reaction time was mid > high > low (Tukey-adjusted $p < 0.05$ and $p < 0.001$, respectively), and (2) in the identification of tones produced by females, reaction time was mid > low > high (Tukey-adjusted $p < 0.001$ and $p < 0.05$, respectively).

Table 3. Summary of the best-fitting linear mixed-effects model for reaction time (*** $p < 0.001$).

Effects	Measures		
	Estimated coefficient	Std. error	*t*-value
(intercept)	6.654	0.034	196.14***
Tone (L)	0.095	0.027	3.56***
Tone (M)	0.263	0.025	10.58***
Tone (H):Gender (M)	0.183	0.028	6.59***
Tone (L):Gender (M)	−0.179	0.026	−6.78***
Tone (M):Gender (M)	−0.002	0.026	−0.06

4. Discussion

With a high-variability testing paradigm that limits the acoustic cues available for speaker normalization, of which previous cross-linguistic tone perception studies had various controls, this study investigated whether the ability to perceive tones relative to each speaker's *F*0 range is modulated by listeners' tone experience. The results showed that Cantonese level tone identification accuracy was subject to linguistic experience (native tonal vs non-native tonal vs non-native non-tonal) as well as the acoustics of tones (speaker gender and tone height). Given the isolated level-tone stimuli mixed across talkers, only native Cantonese listeners could reliably identify tones of all speaker gender*tone height combinations above the chance level. Mandarin listeners had comparable performance to native controls, except in the perception of high-level tones produced by male speakers and low-level tones produced by female speakers (which were misidentified as mid-level tones almost half of the time). More importantly, Mandarin listeners showed a higher accuracy than English listeners almost across the board, except in the perception of low-tones produced by males and high-tones produced by females, where both groups did equally well. It is therefore suggested that Mandarin listeners' greater ability than English listeners to normalize for inter-speaker variation, in the absence of acoustic cues typically used for speaker normalization, in perceiving a non-native level-tone contrast may be attributed to the two groups' different tone language experience. That is, tone listeners' experience with processing *F*0 variations at the lexical level may be transferred to aid naive perception of tonal contrasts in another language.

This result contrasts with findings from Qin and Mok (2011) and Qin and Jongman (2015) and the pretest of Francis *et al.* (2008), in so far as Cantonese level tone perception is concerned. It may be that, as in Qin and Mok (2011) and Qin and Jongman (2015), English listeners' advantage of perceptual sensitivity to *F*0 height is only present in a more low-level psychoacoustic task (i.e., an AX discrimination task) with single-speaker stimuli (such that speaker normalization is not required). In an identification task that requires linguistic processing of *F*0 variation (as used in the current study and Francis *et al.*, 2008), English listeners' sensitivity to subtle differences among level tones alone does not guarantee better performance than Mandarin listeners. Additionally, the discrepant findings between Francis *et al.* (2008) (null differences between the two groups) and the current findings could be attributed to one major methodological difference—the stimuli were produced by one speaker, which requires no speaker normalization in identification, in the pretest of the study by Francis *et al.* (2008). As a result, their task may not be demanding enough to differentiate the two groups (i.e., both groups' accuracy rates in level tone identifications were noticeably higher than those reported in the current study). This possibility can be further tested by replicating studies that used single-speaker stimuli and reported null tone language experience effects (e.g., Hao, 2012; Lee *et al.*, 1996) with a high-variability listening paradigm.

Note that our Mandarin listeners were from Taiwan, and 16 of them had various proficiency levels in Taiwanese. On a five-point scale (1: poor; 5: native), 8 Taiwanese-speaking listeners rated 5 points on their Taiwanese proficiency, two rated 4 points and six rated 3 points. Since there are two level tones in Taiwanese that contrast exclusively in pitch height, it is possible that Mandarin listeners who spoke Taiwanese would be better at the current task than monolingual Mandarin listeners, such that the tone language experience advantage may not actually be present upon comparison of monolingual Mandarin and English listeners' performance. Our data does not seem to support this speculation, as listeners that reported to not speak Taiwanese had an average accuracy of 54.8%, while those who rated 5 points on their Taiwanese proficiency showed an accuracy of 49.9%. The results of a mixed-effects analysis on Mandarin group's identification accuracy, however, did not confirm a significant effect of self-rated Taiwanese proficiency on Cantonese level tone identification. Additionally, a previous study by Chiao *et al.* (2011) actually found Taiwanese speakers to have more difficulties perceiving a four-way level tonal contrast in Toura than speakers of Vietnamese, which only has one level tone. Taken together, we suspect that even if monolingual Mandarin listeners had been recruited for this study, beneficial effects of tone language experience would still be observed.

Besides the cross-group accuracy differences, we also observed similar perception patterns among groups: All three groups had most accurate performance in identifying low-tones produced by male speakers and high-tones produced by female speakers. That is, our listeners, regardless of their linguistic background, were able to most accurately identify tone heights that were at the extremes of the speakers' *F*0 range—in agreement with the findings of Lee *et al.* (2014) of Taiwanese level tones identified by Taiwanese and English listeners. Tones produced in the middle of the speakers' *F*0 range, which native Cantonese listeners could all identify above chance, produced a lower accuracy. The reaction time data, though unable to discriminate the three groups (which was the case in Lee *et al.* as well, and which we cannot speculate further without evidence of an accuracy-speed tradeoff), concurred with the accuracy analysis in regards to the effects of the acoustics of tones: tones produced at the extremes of the speakers' $F0$ range were identified faster than those in between.

In conclusion, this study provides evidence for the beneficial effects of tone language experience on the perception of non-native level-tone contrasts under the high-variability listening condition. Whether the cross-group accuracy differences (as a result of language experience) and similarities (due to psychoacoustic factors) observed using a similar experimental paradigm are also present in the perception of contour tones calls for future research.

Acknowledgments

This work was supported by Ministry of Science and Technology, Taiwan, Republic of China under Grant No. MOST 104-2410-H-027-008.

References and links

[1]Given that the study by Francis *et al.* (2008) is a training study, we only referred to the results of their pre-test. It should be noted that Lee *et al.* (1996) also reported comparable performance between Mandarin listeners' and English listeners' perception of Cantonese tones. However, Lee *et al.* did not provide further information as to whether the two groups were equally good at discriminating all Cantonese tone pairs, which makes it impossible to compare their results to those of Francis *et al.* (2008) and Qin and Mok (2011) and Qin and Jongman (2015).

[2]One reviewer asked if native listeners' overall low performance could be attributed to the merger between mid- and low-tones. According to Mok *et al.* (2013), tone-merging participants could still distinguish the tones as accurately as non-merging participants in perception. In addition, Wong and Diehl's (2003) native Cantonese listeners performed at an accuracy of 48.6% in an experimental condition comparable to ours. In this regard, our native controls' performance was not particularly low. However, we acknowledge that a tone perception screening task is needed if listeners with a (near) tone merger are to be excluded.

[3]The results did not reveal a significant block effect, suggesting that identification accuracy was not affected by potentially different degrees of tonal overlap among blocks. One reviewer questioned if Mandarin listeners, who have more experience with inter-speaker variability in tones than English listeners, showed a performance improvement over blocks under the high-variability listening condition. The results suggest that identification accuracy did not differ over blocks for any listener group.

Chiao, W.-H., Kabak, B., and Braun, B. (**2011**). "When more is less: Non-native perception of level tone contrasts," in *Proceedings of Psycholinguistic Representation of Tone Conference*, Hong Kong, China.

Francis, A. L., Ciocca, V., Ma, L., and Fenn, K. (**2008**). "Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers," J. Phonetics **36**, 268–294.

Gandour, J. T. (**1983**). "Tone perception in Far Eastern languages," J. Phonetics **11**(2), 149–175.

Gandour, J. T., and Harshman, R. A. (**1978**). "Cross language differences in tone perception: A multidimensional scaling investigation," Lang. Speech **21**(1), 1–33.

Hallé, P. A., Chang, Y. C., and Best, C. T. (**2004**). "Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners," J. Phonetics **32**, 395–421.

Hao, Y.-C. (**2012**). "Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers," J. Phonetics **40**(2), 269–279.

Lee, C.-Y., Lee, Y.-F., and Shr, C.-L. (**2011**). "Perception of musical and lexical tones by Taiwanese-speaking musicians," J. Acoust. Soc. Am. **130**, 526–535.

Lee, C.-Y., Lekich, A., and Zhang, Y. (**2014**). "Perception of pitch height in lexical and musical tones by English-speaking musicians and nonmusicians," J. Acoust. Soc. Am. **135**, 1607–1615.

Lee, C.-Y., Tao, L., and Bond, Z. S. (**2009**). "Speaker variability and context in the identification of fragmented Mandarin tones by native and non-native listeners," J. Phonetics **37**(1), 1–15.

Lee, Y.-S., Vakoch, D. A., and Wurm, L. H. (**1996**). "Tone perception in Cantonese and Mandarin: A cross-linguistic comparison," J. Psycholing. Res. **25**(5), 527–542.

Liang, J., and van Heuven, V. J. (**2007**). "Chinese tone and intonation perceived by L1and L2 listeners," in *Tones and Tunes*, edited by C. Gussenhoven and T. Riad, Vol. 2: Experimental studies in word and sentence prosody (Mouton de Gruyter, Berlin, New York), pp. 27–61.

Mok, P. P., Zuo, D., and Wong, P. W. (**2013**). "Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese," Lang. Variat. Change **25**(3), 341–370.

Qin, Z., and Jongman, A. (**2015**). "Does second language experience modulate perception of tones in a third language?," Lang. Speech **59**(3), 318–338.

Qin, Z., and Mok, P. (**2011**). "Perception of Cantonese tones by Mandarin, English and French speakers," in *Proceedings of the ICPhS XVII*, Hong Kong, China.

So, C. K., and Best, C. T. (**2010**). "Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences," Lang. Speech **53**(2), 273–293.

tonometric.com (Last viewed August 30, 2016).

Wayland, R. P., and Guion, S. G. (**2004**). "Training English and Chinese listeners to perceive Thai tones: A preliminary report," Lang. Learn. **54**(4), 681–712.

Wong, P. C., and Diehl, R. L. (**2003**). "Perceptual normalization for inter- and intratalker variation in Cantonese level tones," J. Speech Lang. Hear. Res **46**(2), 413–421.