The separation between music and speech: Evidence from the perception of Cantonese tones[a)]

P. K. Peggy Mok[b)] and Donghui Zuo
Department of Linguistics and Modern Languages, Leung Kau Kui Building, Chinese University of Hong Kong, Shatin, Hong Kong

(Received 30 April 2012; revised 18 July 2012; accepted 30 July 2012)

This study investigates the relationship between music and speech, testing whether musical training has any facilitatory effects on native tone language speakers. Some Cantonese tone pairs are merging in recent years. The merging subjects have poorer general lexical tone perception than the control subjects. Previous studies showed that musical training facilitates lexical tone perception of nontone language speakers, but it is unclear if the same is true for tone language speakers. Three groups of listeners (standard Cantonese, merging Cantonese, nontone) with and without musical training participated in AX discrimination tasks of Cantonese monosyllables and pure tones resynthesized from Cantonese lexical tones. Results show that while musical training enhances lexical tone perception of nontone listeners, it has little influence on Cantonese listeners. The findings suggest that the linguistic use of tones is more fundamental and more robust than musical tones. Our results are compatible with the idea that linguistic and musical mechanisms belong to separate but overlapping domains.
 [http://dx.doi.org/10.1121/1.4747010]

PACS number(s): 43.75.Cd [DD] Pages: 2711–2720

I. INTRODUCTION

This study investigates the relationship between musical training and the perception of lexical tones by comparing the perception of Cantonese tone pairs by native Cantonese and nontone listeners with and without musical training. Its specific focus is whether musical training has any facilitatory effect on native Cantonese listeners who are merging lexical tones.

Much recent research interest is drawn into the relationship between musical training and lexical tone perception. Most previous studies on the relationship between music and speech either compared tone perception by nontone language speakers who are musicians and non-musicians or compared these nontone speakers with tone-language speakers with no musical training (e.g., Wong *et al*., 2007; Lee and Hung, 2008; Chandrasekaran *et al*., 2009; Bidelman *et al*., 2010, 2011). Previous studies have shown that musical training can facilitate lexical tone perception of nontone language speakers (e.g., Gottfried, 2007; Lee and Hung, 2008; Wayland *et al*., 2010). In addition, there are significantly more musicians speaking Mandarin (a tone language) with absolute pitch than musicians speaking American English; and the prevalence of absolute pitch is negatively correlated with the age of onset of musical training (Deutsch *et al*., 2006; Deutsch *et al*., 2009; Lee *et al*., 2011). These results suggest that similar processing mechanisms may be involved in lexical and musical tone perception even if the nature of such mechanisms is still unclear (Schellenberg and Peretz, 2008). However, a crucial question regarding the "same mechanisms" proposal remains unanswered: Whether musical training can also facilitate lexical tone perception of native tone language speakers as well. Such evidence is understandably elusive because native speakers of tone languages are believed to perform at ceiling and are used as reference for comparison. A tone language with acoustically similar tone pairs that are undergoing tone change offers the best opportunity to investigate this possibility. Cantonese is a case in point.

Hong Kong Cantonese is well-known for its complex tone system with six lexical tones (T): T1 (high-level [55]), T2 (high-rising [25]), T3 (mid-level [33]), T4 (low-falling [21]), T5 (low-rising [23]), and T6 (low-level [22]). There are three tones for checked syllables ending with final stop consonants (-p, -t, -k): T7 (high-stopped [5]), T8 (mid-stopped [3]), and T9 (low-stopped [2]), which are considered allotones of T1, T3, and T6, respectively (Bauer and Benedict, 1997). In total, then, the Cantonese tone inventory contains three level tones (T1 [55], T3 [33], T6 [22]), two rising tones (T2 [25], T5 [23]), and one falling tone (T4 [21]). The Cantonese "tone space" is very crowded because apart from T1 [55], the remaining five tones are all situated in the low to mid pitch range, and four of them share the same starting pitch height [2]. Several tone pairs, T2 [25]/T5 [23], T3 [33]/T6 [22], T4 [21]/T6 [22], are acoustically very similar. They are confusable even for adult native speakers (Varley and So, 1995).

In recent years, some Cantonese speakers in Hong Kong no longer clearly distinguish all six tones in their production. The most notable merging pair is T2 [25] vs T5 [23] (Bauer *et al*., 2003). Some speakers also have a tendency to merge T3 [33] and T6 [22] and T4 [21] and T6 [22], respectively (Mok and Wong, 2010a, 2010b). Tone merging is a relatively recent phenomenon in Hong Kong. The acoustic similarity

[a)]Portion of this work was presented in "Effects of tone merging and musical training on Cantonese tone perception," Speech Prosody 2012, Shanghai, China, May, 2012.
[b)]Author to whom correspondence should be addressed. Electronic mail: peggymok@cuhk.edu.hk

between these tone pairs renders them particularly susceptible to sound change. Sociolinguistically, the influx of migrants from China, language contact between Cantonese and other languages (other Chinese varieties), and the increasing influence of Mandarin are likely factors contributing to the merging of tones in Hong Kong. Mok and Wong's (2010a) preliminary data showed that the merging native speakers have poorer general tone perception of Cantonese tones than their non-merging native counterparts do; this is not limited only to the merging tone pairs. This implies that the merging speakers are less sensitive to linguistic tone differences even though they are also native speakers. This situation raises the interesting question of whether musical training has any facilitatory effect on tone perception of these merging native speakers.

Few studies have investigated the effect of musical training on native tone language speakers. Deutsch and colleagues (Deutsch, 2002; Deutsch *et al.*, 2004; Deutsch *et al.*, 2006; Deutsch *et al.*, 2009) proposed a concrete hypothesis that the rare musical ability of absolute pitch originally evolved to subserve speech and that processing of lexical and musical pitch share the same neural circuitry. They argued that musical tones and lexical tones are treated in the same way by the brain. Infants born into a tone language environment develop the neural circuitry for processing tones (both lexical and musical). When they later begin music lessons, the neural circuitry for tones is already in place, and therefore they should have an advantage in processing musical tones compared to nontone language speakers. As mentioned in the preceding text, Deutsch and colleagues have shown that there are significantly more musicians speaking a tone language with absolute pitch than musicians speaking English; and the prevalence of absolute pitch is negatively correlated with the age of onset of musical training.

Other studies on the relationship between music and speech have not advanced such specific hypothesis for tone language speakers. Although Deutsch's hypothesis provides concrete predictions about music and speech regarding absolute pitch, there are still many unknowns about the more general effects of musical training on tone language speakers as there are substantial differences in how tone patterns are used in music and language. Besides absolute pitch is unnecessary for the development of normal musical ability (e.g., see relevant discussion in Patel, 2008; Bidelman *et al.*, 2010, 2011). Deutsch *et al.* (2004, p. 340) themselves also pointed out that absolute pitch is not necessarily accompanied by superior performance on other musical processing tasks. Therefore more studies are needed to investigate the domain-(non)specificity of music and speech for tone language speakers.

The current study compares tone perception of merging Cantonese speakers ("merging Cantonese" hereafter) with that of non-merging Cantonese speakers, who clearly distinguish all six Cantonese tones ("standard Cantonese" hereafter). Additionally, it investigates the relationship between lexical tone perception and musical training by comparing the data of both Cantonese and nontone listeners with advanced musical training ("musicians" hereafter) and those with no or very limited musical experience ("non-musicians" hereafter). Cantonese is an interesting target language because of its complex tone system as described in the preceding text. Previous studies on similar topics used either the Mandarin four-tone system (i.e., level, rising, dipping, and falling) (e.g., Lee and Hung, 2008) or the Assamese two-tone system (i.e., falling and rising) (Wayland *et al.*, 2010) in which the major tone difference lies in pitch direction. In these studies, musicians consistently outperformed non-musicians; this suggests that musical training facilitates listeners' ability to identify non-native lexical tones. However, it is unclear whether the same musical facilitatory effects can also be found when the nontone listeners need to distinguish much more subtle differences found in Cantonese in which both pitch height and the magnitude of change are important cues for tonal distinction (Gandour, 1981).

The core research question of the current study is whether musical training has any facilitatory effects on merging Cantonese speakers. First of all, are there Cantonese musicians who are merging lexical tones? If so, would their musical training have any bearing on the degree of tone merge? Specifically, would the degree of tone merge for merging Cantonese musicians be less severe than that of merging Cantonese non-musicians? Would the same musical training effects be found in both production and perception? These interesting issues on musical training and native tone language speakers have not been discussed in the literature before. Elucidation of these questions can provide greater understanding of the intricate relationship between music and speech.

II. METHOD

A. Subjects

Three groups of subjects participated in the experiment. The first group consisted of 34 native speakers of English or French who had not learned any tone languages before (nontone). The second group consisted of 30 Cantonese speakers who clearly distinguished all six tones in their production (standard Cantonese), and the third group consisted of 28 Cantonese speakers who did not clearly distinguish all six tones in their production (merging Cantonese).

The merging Cantonese subjects were selected from 169 native speakers with a screening process. Each speaker was recorded reading a list of 30 words (five different words $\times$ six tones) embedded in a short carrier phrase. The recordings of all 169 speakers were auditorily checked independently by two native speakers of Cantonese with phonetic training who clearly distinguish all six tones (the first author included) to determine whether the speaker was likely to merge the tones. Only those who were identified as merging Cantonese by both judges were included in this study. The 30 standard Cantonese subjects were also recruited from this pool of 169 speakers. Their production was confirmed by the two native judges as clearly distinguishing the tones. Acoustic analyses of the tones produced by the two subject groups were conducted to corroborate the division of these two groups of subjects as observed in the impressionistic screening. Details are given in Sec. III.

Although the standard Cantonese and merging Cantonese subjects were so divided based only on their speech

TABLE I. Linguistic backgrounds of the standard Cantonese and merging Cantonese subjects.

	No. of subjects contacted	Born in Mainland China	Own language use at home: Cantonese only	Own language use at home: Cantonese and other languages	Parents' language use at home: Cantonese only	Parents' language use at home: Cantonese and other languages	Mean onset age of Mandarin training	Mean length of Mandarin training (yr)
Standard	29	3 (10.3%)	27 (93.1%)	2 (6.9%)	17 (58.6%)	12 (41.4%)	7.1 (S.D. 2.1)	7.4 (S.D. 2.7)
Merging	24	4 (16.7%)	23 (95.8%)	1 (4.2%)	14 (58.3%)	10 (41.7%)	7.8 (S.D. 2.2)	7.5 (S.D. 2.7)

production, we also tracked them down for further detail of their linguistic backgrounds. We could contact more than 90% of them finally (29 standard and 24 merging subjects). Table I shows that the two groups of subjects have quite similar linguistic backgrounds according to our questions. Except two merging Cantonese subjects who came to Hong Kong from Mainland China at around age 10, all other subjects grew up in Hong Kong either from birth or from a young age (before age 5) and spoke Cantonese as their home language. Some subjects in both groups were also exposed to other languages (mainly Chinese dialects) at home. They attended local schools and began learning English as a second language when they entered kindergarten (around age 3) or primary one (around age 6) for those who arrived in Hong Kong later. All subjects passed the Use of English in the Hong Kong Certificate of Education Examination (HKCEE) with varying grades. Some subjects were unwilling to disclose their exact grades. For those who reported, there were 7 As, 9 Bs, 9 Cs, and 4 Ds in the standard Cantonese group, and 5 As, 7 Bs, 6 Cs, and 1 D in the merging Cantonese group. The subjects also learned Mandarin in school. As Mandarin was not a compulsory subject in HKCEE, there is no objective measure of their Mandarin proficiency. Nevertheless, both groups of subjects had similar length of training in Mandarin, and all reported that they had intermediate to advanced proficiency in this language.

The three groups of subjects were further divided into three categories by their musical background. Subjects with more than 7 years of formal musical training in any instrument or vocal singing and had played music regularly in the past 2 years were classified as musicians. Subjects with no more than 2 years of casual musical experience and had not played music regularly in the past 2 years were classified as non-musicians. Subjects with musical experience between these two extremes were classified as intermediate. Intermediate subjects were excluded from the analysis of musical training as their musical background is too ambiguous. The numbers of subjects in each group can be found in Table II.

TABLE II. Numbers of subjects with different language and music backgrounds.

	Nontone	Standard Cantonese	Merging Cantonese
Musician	10	10	10
Intermediate	10	10	7
Non-musician	14	10	11
Total	34	30	28

All the subjects were students (either local or exchange) at the Chinese University of Hong Kong except five nontone musicians who were recruited at the University of Chicago in the United States.

B. Materials

The experiment consisted of two AX discrimination tasks of Cantonese monosyllables and pure tones resynthesized from the six Cantonese tones. The monosyllable stimuli were produced by a female researcher and speech therapist, who clearly distinguishes the six Cantonese tones in her production. The tokens were paired up to form two types of stimuli pairs–120 AA pairs (same-tone pairs) and 120 AB pairs (different-tone pairs). Altogether 60 target monosyllables with open or nasal endings (six tones $\times$ 10 syllables) were chosen as the AA pairs together with 60 dummy items to balance the number of the AB pairs. Dummy items were the checked syllables with final stop consonants (-p, -t, -k) which are considered allotones of T1, T3, and T6. These dummy items were excluded from analysis. For the AB pairs, two syllables of each tone that also appeared in the AA pairs were chosen. These two syllables were paired with the other five tones to form the AB pairs, for example, T1/T2, T1/T3, T1/T4, T1/T5, T1/T6. The order of the AB pairs was counter-balanced, i.e., both AB and BA pairs were included. This resulted in 120 AB pairs (six tones $\times$ two syllables $\times$ five matching tones $\times$ two orders). The 120 AA and 120 AB pairs were randomized in the perception experiment.

The six pure tone stimuli were resynthesized from a Cantonese monosyllable [wai] produced by the same female researcher mentioned above using PRAAT (4.6.29). The six tones with [wai] are all meaningful Cantonese monosyllables. The pitch contours of the resynthesized pure tone stimuli were exactly the same as the six Cantonese lexical tones, but all segmental information was removed. There are in total 60 AA pairs and 60 AB pairs (six tones $\times$ five matching tones $\times$ two orders). The pure tone tokens were also randomized in the experiment.

C. Procedures

The subjects participated in the perception experiment individually in quiet rooms. They performed the monosyllable task before the pure tone task. The stimuli were presented to them through a stereo headphone using E-PRIME 2.0 PROFESSIONAL with a desktop computer. A short practice before each task and short breaks during the tasks were given. The subjects were asked to indicate whether the two tokens in the stimuli pairs carry the same tone or not by pressing different buttons on a serial response box. They were encouraged to respond as accurately and as quickly as possible. Both accuracy and reaction time were collected. The inter-stimulus interval (ISI) was 500 ms. No feedback was given. All reaction time values longer than 3000 ms were counted as missing responses and were excluded from analysis. This resulted in 0.6% loss of data.

III. RESULTS

A. Production

The focus of this study is on tone perception. Nevertheless, acoustic analyses of the tones produced by the merging Cantonese and standard Cantonese subjects in the screening test were conducted to confirm the division of these two groups of subjects based on impressionistic screening. These production data can also reveal the patterns of tone merging in the subjects. Previous studies have shown that Cantonese tones are primarily discriminated by their pitch contours especially in the later portion of the tones. Other features like intensity and duration are not important (Vance, 1977; Khouw and Ciocca, 2007). Therefore only pitch was analyzed in the present study. The beginning and ending of each tone were marked manually, and the F0 was measured at every +10% from the beginning of each tone. The F0 values at the 5th (50%) and 9th (90%) deciles represent the middle and the offset of the tone. Discriminant analysis was used to calculate the classification rate of the tones for each speaker based on these two predictors (50% point and 90% point).

In addition to capturing the later pitch contours with two time points (50% and 90%) mentioned in the previous paragraph, there is another reason why only two predictors were used for discriminant analysis. Due to creakiness in the low tones, particularly T4, F0 could not be measured from some tokens. This resulted in fewer than five usable tokens for a particular tone for some subjects. Discriminant analysis requires the sample size of the smallest group to be larger than the number of predictors, i.e., the number of tokens in each tone group should be larger than the number of predictors. Therefore only two predictors were used to include as many speakers as possible. Finally, 47 (23 standard vs 24 merging) of the 58 recruited speakers were included in the analysis.

T1[55] was excluded from the discriminant analysis as it is well separated from other tones and is not involved in any merging pairs. Each of the five remaining tones was treated as a group of five members (i.e., the five tokens produced in the screening test). Leave-one-out classification, in which each token is classified by the functions derived from all the other tokens, was conducted to predict the group membership of each token. More specifically, in the classification process, each token was treated as an unknown sample, and all the other tokens were used to generate a model for differentiating the tones that was then used to predict the group membership of that unknown sample. A higher classification rate indicates better separation of tones, so the standard Cantonese subjects should have higher classification rates than the merging Cantonese subjects. The classification rate of all the 47 speakers (mean = 70.2%, S.D. = 12.09%) is much higher than the chance level (1/5 = 20%); this suggests that the 50% and 90% points can already discriminate different tones very well.

An independent t-test shows that the classification rate of the standard Cantonese group (mean = 77.58, S.D. = 8.46) is significantly higher than the merging Cantonese group (mean = 63.25, S.D. = 10.91) [$t(45) = 5.016$, $P < 0.0001$]. The acoustic analysis confirms the separation of the standard Cantonese and merging Cantonese groups by the native judges and also shows that the tones produced by the merging Cantonese subjects were more similar than those produced by the standard Cantonese subjects.

In addition, the pitch height of each tone was examined as there are multiple level and contour tones in Cantonese. For comparison consistency, the F0 values at the 90% points (i.e., the tone offset) for all the six tones were used as the 50% point is not suitable for distinguishing the contour tones because the major differences between them are at the end of the tones. The left panel of Table III shows that although there is no obvious difference in the pitch values of individual tones between the male subjects, the pitch values of the female standard Cantonese subjects are consistently higher than those of the female merging Cantonese subjects. The difference is significant for T1 [$t(35) = 2.206$, $P = 0.034$] and approaching significance for T2 [$t(35) = 2.018$, $P = 0.051$] and T3 [$t(35) = 1.865$, $P = 0.071$].

The differences between the merging tone pairs, and between the high level T1 [55] and the low level T6 [22]

TABLE III. Mean pitch values in hertz (standard deviations) of each Cantonese tone at the 90% point and the mean quotients (standard deviations) of several tone pairs for tone space estimation.

Subjects		Total number	Individual tones						Quotients for tone space estimation			
			T1	T2	T3	T4	T5	T6	T1/T6	T3/T6	T2/T5	T4/T6
Female	Standard	17	257 (24.3)	229 (27.5)	204 (19.8)	169 (15.3)	204 (19.7)	188 (18.1)	1.368 (0.07)	1.085 (0.04)	1.125 (0.06)	1.115 (0.04)
	Merging	20	239 (25.8)	212 (24.4)	192 (18.9)	163 (16.2)	195 (20.3)	179 (13.5)	1.331 (0.08)	1.071 (0.04)	1.086 (0.05)	1.105 (0.10)
Male	Standard	7	135 (19.9)	126 (18.9)	110 (16.4)	94 (13.2)	109 (14.6)	101 (14.2)	1.334 (0.06)	1.092 (0.04)	1.157 (0.03)	1.082 (0.04)
	Merging	6	134 (27.6)	120 (19.9)	111 (19.2)	94 (13.8)	115 (17.4)	105 (18.7)	1.280 (0.11)	1.058 (0.01)	1.048 (0.05)	1.108 (0.05)

TABLE IV. The mean quotients (standard deviations) of several tone pairs for tone space estimation in standard and merging Cantonese groups.

Subjects		Quotients for tone space estimation			
		T1/T6	T3/T6	T2/T5	T4/T6
Standard	Musicians	1.375 (0.67)	1.083 (0.03)	1.141 (0.06)	1.125 (0.04)
	Non-musicians	1.377 (0.40)	1.099 (0.04)	1.135 (0.04)	1.081 (0.04)
Merging	Musicians	1.289 (0.10)	1.062 (0.03)	1.076 (0.05)	1.089 (0.05)
	Non-musicians	1.346 (0.09)	1.069 (0.04)	1.095 (0.06)	1.089 (0.10)

were calculated to estimate the size of the "tone space." The pitch values of the higher tones (numerator) were divided by those of the lower tones (denominator). The resultant quotients are thus normalized for individual differences. Higher quotients indicate that the tones are more dispersed. The right panel of Table III shows that except for the quotients of T4/T6 for male subjects, the standard Cantonese subjects had higher quotients than the merging Cantonese subjects. The difference is significant for the T2/T5 pair for both male [$t(11)=5.072$, $P<0.0001$] and female subjects [$t(35)=2.091$, $P=0.044$]. If we combine the quotients of male and female subjects (as the quotients are normalized), in addition to the T2/T5 pair, the difference between T3/T6 also approaches significance [$t(48)=1.845$, $P=0.071$]. The lack of statistical power due to limited data (only five tokens for each tone) can explain the nonsignificance of some comparisons. Nevertheless, the production data clearly suggest that the lexical tones were more distinct for the standard Cantonese than the merging Cantonese subjects, echoing the conclusion based on discriminant analysis in the preceding text.

Finally, it is worth checking whether there is any production difference between musicians and non-musicians within the standard Cantonese group and the merging Cantonese group, respectively. The mean classification rates in discriminant analysis for non-musicians were even higher than those for the musicians in both groups, although the difference was not significant: Standard Cantonese [musicians: 75.22 (S.D. 7.8) vs non-musicians: 82.09 (S.D. 7.7), $t(15)=-1.793$, $P=0.093$]; and merging Cantonese [musicians: 57.36 (S.D. 13.8) vs non-musicians 64.44 (S.D. 7.6), $t(17)=-1.411$, $P=0.176$]. In terms of the size of the tone space, Table IV shows that there is no consistent pattern between the quotients for musicians and non-musicians in both groups. Among all comparisons between musicians and non-musicians, only the difference for T4/T6 in the standard Cantonese group is significant [$t(15)=2.204$, $P=0.044$]. Therefore we can conclude that musical training did not affect tone production of both standard and merging Cantonese subjects.

B. Perception

The accuracy (%Correct) and the reaction time data of the three subject groups were first analyzed regardless of musical training. We log-transformed the reaction time data (LogRT) for normalization. Figure 1 shows the averaged %Correct and LogRT of the three subject groups in the monosyllable and pure tone tasks. In general, the standard Cantonese subjects were the quickest and the most accurate while the nontone subjects were the slowest and the least accurate. One-way ANOVAs confirm that both the %Correct [$F(2,\ 89)=7.727$, $P=0.001$] and the LogRT [$F(2,\ 89)=11.980$, $P<0.0001$] are significantly different between groups in the monosyllable task [Fig. 1(A)]. *Post hoc* tests with Bonferroni corrections indicate that both the standard Cantonese ($P=0.001$) and merging Cantonese ($P=0.011$) subjects were significantly more accurate than the nontone subjects, but there is no significant difference between the standard and merging subjects. The standard Cantonese subjects were significantly faster in terms of LogRT than both the merging Cantonese ($P=0.022$) and nontone ($P<0.001$) subjects, while there is no significant difference between the merging Cantonese and nontone groups ($P=0.157$). In summary, the merging Cantonese subjects resemble the standard Cantonese subjects in accuracy but resemble the nontone subjects in reaction time for monosyllables.

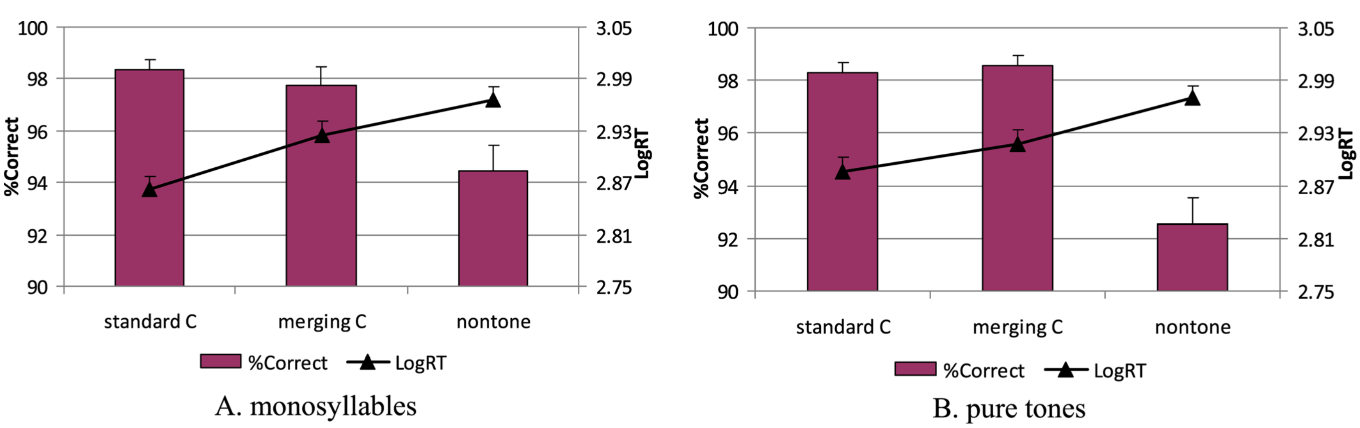

FIG. 1. (Color online) Accuracy (%Correct) and log-transformed reaction time (LogRT) of the three subject groups in the (A) monosyllable task and (B) pure tone task. The error bars show one standard error.

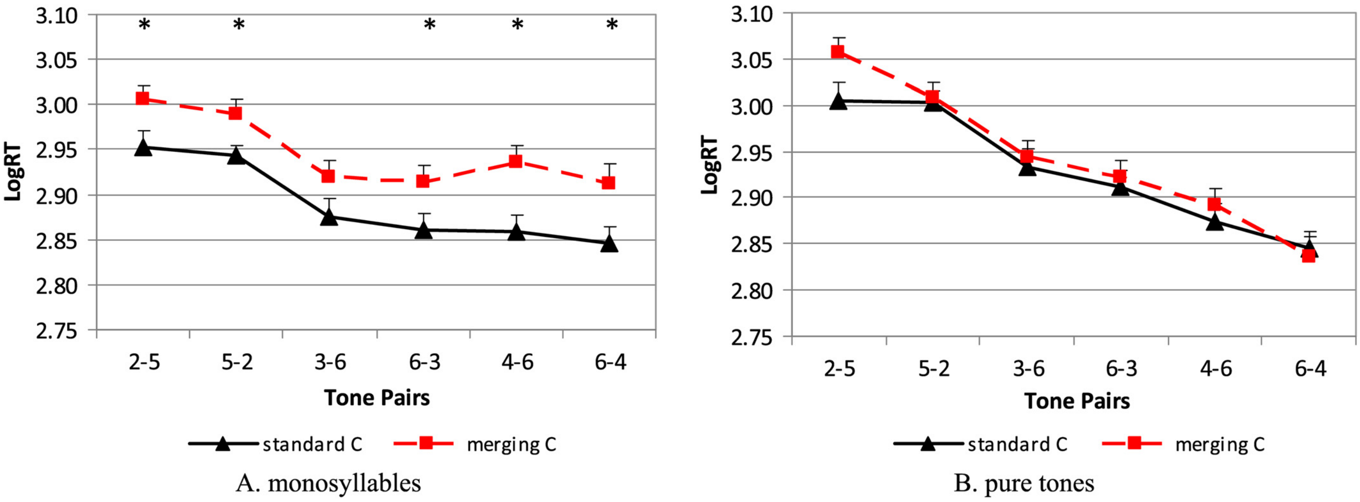

FIG. 2. (Color online) LogRT of the standard Cantonese and merging Cantonese subjects for the confusing tone pairs in (A) monosyllable task and (B) pure tone task. The error bars show one standard error. Significant comparisons ($P < 0.05$) are marked with an asterisk.

Although the patterns look similar, there is an interesting language difference in the pure tone task [Fig. 1(B)]. Again, both the %Correct [$F(2, 89) = 26.843$, $P < 0.0001$] and the LogRT [$F(2, 89) = 8.791$, $P < 0.0001$] are significantly different between groups. *Post hoc* tests with Bonferroni correction indicate that the standard Cantonese and merging Cantonese subjects are not significantly different in either %Corret ($P = 1$) or LogRT ($P = 0.408$), while both groups are significantly better than the nontone subjects in %Correct (standard: $P < 0.0001$; merging: $P < 0.0001$) and LogRT (standard: $P < 0.0001$; merging: $P = 0.042$).

To further investigate the interesting difference in LogRT between the monosyllable and pure tone tasks for the standard Cantonese and merging Cantonese subjects, we compared the confusing tone pairs (T2/T5, T3/T6, T4/T6) individually as these are the pairs merging in Hong Kong Cantonese in recent years (Bauer *et al.*, 2003; Mok and Wong, 2010a, 2010b). The standard subjects were significantly faster than the merging subjects for most confusing tone pairs in the monosyllable task [see Fig. 2(A), T2/T5 $P = 0.05$; T5/T2 $P = 0.014$; T3/T6 $P = 0.105$, T6/T3 $P = 0.032$, T4/T6 $P = 0.003$, T6/T4 $P = 0.027$], while there is no significant difference between them in the pure tone task [Fig. 2(B)], which suggests that the standard Cantonese and the merging Cantonese subjects were equally sensitive to pure tones. In other words, even though the pure tone stimuli were resynthesized from Cantonese monosyllables, and the pitch contours were the same in both tasks, there was a difference in the subjects' responses. From these results, it is clear that when linguistic information is filtered out, the standard Cantonese subjects do not have any advantage over the merging Cantonese subjects in perceiving pitch contours identical to the canonical lexical tones.

The main research question of the present study is whether musical training can facilitate tone perception of the merging Cantonese subjects, who resembled nontone subjects in reaction time and were less sensitive to lexical tones than the standard Cantonese subjects were. The musicians' and non-musicians' %Correct and LogRT data in the three subject groups were compared. Subjects with intermediate musical training were excluded for a clearer picture (see Table I). The results in Table V confirm that, similar to previous studies, the nontone musicians (%Correct = 98.35) are significantly more accurate than the nontone non-musicians (%Correct = 91.63) [$t(13.828) = 14.402$, $P = 0.006$] in the monosyllable task, but the differences in reaction time are not significant. It is interesting to note that musical training seems to facilitate the accuracy of nontone musicians to the same level as native speakers (around 98%) [nontone musicians: 98.35 (S. D. = 1.1), standard Cantonese subjects: 98.36 (S. D. = 1.9), merging Cantonese subjects: 97.73 (S. D. = 3.8), $F(2, 65) = 0.397$, $P = 0.674$]. For the native Cantonese subjects (both standard and merging), however, musical training affects neither the %Correct nor the LogRT. Although the merging Cantonese musicians have a higher

TABLE V. Average %Correct and LogRT (standard deviations) of musicians and non-musicians in the monosyllable and pure tone tasks.

			Standard Cantonese	Merging Cantonese	Nontone
Monosyllables	%Correct	Musicians	98.86 (0.92)	98.90 (1.59)	98.35 (1.14)
		Non-musicians	98.53 (2.95)	96.03 (5.68)	91.63 (7.54)
	LogRT	Musicians	2.87 (0.05)	2.92 (0.05)	29.56 (0.09)
		Non-musicians	2.85 (0.09)	2.93 (0.12)	29.64 (0.07)
Pure tones	%Correct	Musicians	97.94 (2.26)	99.23 (0.80)	96.28 (2.83)
		Non-musicians	98.44 (2.33)	97.95 (2.20)	88.93 (6.19)
	LogRT	Musicians	2.92 (0.06)	2.90 (0.07)	29.54 (0.07)
		Non-musicians	2.86 (0.09)	2.93 (0.09)	29.78 (0.09)

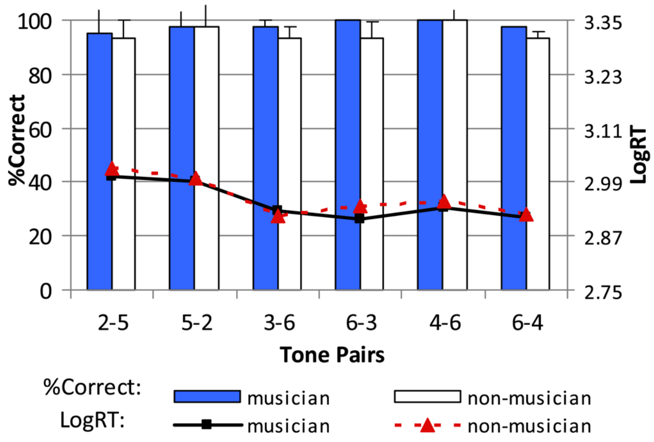

FIG. 3. (Color online) Accuracy (%Correct) and LogRT of the confusing tone pairs in the monosyllable task for the merging Cantonese musicians and merging Cantonese non-musicians.

average %Correct than the merging Cantonese non-musicians (98.9 vs 96.0), the difference is not significant [$t(19) = 1.539$, $P = 0.140$].

The same patterns can be found in the pure tone task. Nontone musicians (%Correct = 96.28) are significantly more accurate than nontone non-musicians (%Correct = 88.93) [$t(19.32) = 4.013$, $P = 0.001$] with no difference in reaction time [$t(22) = -0.240$, $P = 0.812$]. Again, musical training affects neither the %Correct nor the LogRT for the standard Cantonese and merging Cantonese groups.

One consistent pattern across subject groups is that non-musicians have a higher variance than musicians as reflected in the larger standard deviations in %Correct and LogRT in both tasks (except LogRT for the nontone group with a very small difference). This indicates that there is more individual variation in non-musicians than musicians.

Finally, it is worth checking whether musical training has any subtle effect on the perception of the confusing lexical tone pairs for the merging Cantonese subjects as these pairs are most vulnerable. Figure 3 shows the %Correct and LogRT of the confusing tone pairs in the monosyllable task for the merging Cantonese musicians and merging Cantonese non-musicians. As we can see, there is hardly any difference at all. Therefore we can conclude that musical training has no effect on the merging Cantonese subjects who had poorer lexical tone perception.

IV. DISCUSSION

The results of the present study confirm previous findings that native speakers of a tone language perform significantly better than speakers of nontone languages in the perception of linguistic tones. It is likely that they perceive tones in different manners (e.g., see Hallé *et al.*, 2004; Wang *et al.*, 2004; Wayland and Guion, 2004; Bent *et al.*, 2006; Burnham and Mattock, 2007). The higher accuracy of the nontone musicians, as compared with the nontone non-musicians, also confirms the results of previous studies that musical training can facilitate the perception of linguistic tone for nontone language speakers. This hypothesis is valid even when the tone system of the target language makes use of not only pitch direction as in previous studies (e.g., Mandarin), but also the more subtle tone differences in Cantonese which involve both pitch height and the magnitude of change. Therefore, the current findings confirm that musical training has facilitatory effects on the linguistic use of tones for nontone language speakers.

The transfer of pitch perception ability from music to language for nontone listeners implies that there is overlap and interaction between the musical and linguistic domains. It is not surprising as both involve pitch as the major correlate, although there are important differences in how pitch is used and structured in each domain (Bidelman *et al.*, 2010, 2011). Bidelman *et al.* (2010) illustrated the overlap and interaction between the two domains by showing that tone language (Mandarin) speakers and (English) musicians exhibited comparable pitch tracking accuracy in the brain stem in perceiving both Mandarin Tone 2 and the musical interval of a major third. This evidence suggests that abilities in pitch encoding can be transferred across domains. However, their findings of domain-specific extraction of pitch features also clearly demonstrated that the two mechanisms are not homogeneous—musicians' brain stems may be more sensitive to pitches that correspond to discrete notes along the musical scale, while tone language speakers' brain stems may be more sensitive to rapid but continuous pitch changes analogous to those used in lexical tones.

Regarding the specific research questions raised in Sec. I, first, we found musicians who are merging lexical tones. Second, musical training did not have any effect on their tone merge in either production or perception. The finding of no facilitatory effect of musical training in either standard Cantonese or merging Cantonese subjects supports the existence of domain-specific modules and circuitries. Merging Cantonese subjects were less sensitive than the standard Cantonese subjects in general lexical tone perception, but even advanced musical training does not have any effect on their perception of lexical tones. It is possible that the ceiling effect has caused the lack of difference in %Correct between the Cantonese musicians and Cantonese non-musicians as it is sufficient for them to depend solely on their linguistic knowledge to discriminate the tones in their native language. One interesting finding of this study is that although the monosyllable task suggests that the standard Cantonese subjects were significantly faster than the merging Cantonese subjects in perceiving lexical tones, their LogRT was as slow as that of the merging subjects in the pure tone task. This implies that the lexical use of tone is more robust and fundamental than the musical use of tone. When linguistic information is removed, the standard Cantonese subjects have no advantage over the merging Cantonese subjects in perceiving pitch contours identical to lexical tones. The different performance suggests that native tone language speakers may be using two different mechanisms to perceive lexical pitch and musical pitch.

The data from the nontone listeners can also give us some side support for the proposal of different mechanisms for music and speech. Both the monosyllable and pure tone stimuli were meaningless to the foreign listeners. However, the monosyllable stimuli were clearly linguistic in nature in that they were produced by a female voice with varying segmental information. Such information alone can already facilitate the nontone listeners' perception of tones. Their %Correct in the monosyllable task is higher than that of the pure tone task [see Fig. 1, $t(33) = -2.705$, $P = 0.011$]. This interesting difference of the nontone listeners is consistent with the suggestion of the more fundamental role of the linguistic domain over the musical domain.

Following the same reasoning, we would expect the monosyllabic stimuli to be even more linguistic in nature to the native Cantonese listeners, but no difference in %Correct was observed between the monosyllable and pure tone tasks for them. A probable reason for this is because they were performing at ceiling for the monosyllable task. The discrimination task in quiet conditions (particularly the AA pairs) was quite easy for native listeners. The production data show that even the merging Cantonese subjects still had six tone categories, although their tones are more similar than those of the standard Cantonese subjects. It is not surprising, then, that they could still correctly distinguish the lexical tones produced clearly by a professional speaker. Similarity with the lexical tones may also account for the high accuracy of the pure tones by the native subjects. Thus the nature of the tasks and the measure of accuracy (%Correct) are not sensitive enough to capture the difference of the two domains for native Cantonese listeners. The sensitivity issue is further supported by the significant difference between the LogRT in monosyllable and pure tone tasks for the standard Cantonese subjects [$t(29) = 3.105$, $P = 0.004$, see Fig. 1], but not in %Correct. More challenging tasks (e.g., presenting the stimuli in noise) or more precise measure (e.g., neuro-imaging) are needed to reveal the subtle differences for the native subjects.

In fact, the different acquisition patterns of language and music also speak for the separation of the two domains. Although some aspects of music can be acquired implicitly (Ettlinger *et al.*, 2011), most musical skills (especially those relating to musicianship) require explicit training and frequent disciplined practice for years. There is also wide individual variation in musical abilities. Language, however, is naturally and implicitly acquired by all children with normal development. No explicit instruction is required for first language acquisition.

Regarding the patterns of acquisition, the hypothesis by Deutsch and colleagues on absolute pitch and lexical tone is relevant here (Deutsch, 2002; Deutsch *et al.*, 2004; Deutsch *et al.*, 2006; Deutsch *et al.*, 2009). As discussed in Sec. I, Deutsch argued that absolute pitch and lexical tone perception share the same neural circuitry and critical age of acquisition. Such general neural circuitry would have been established in infancy for tone language speakers. Her hypothesis predicts that musical training should not affect lexical tone perception of Cantonese native speakers because early linguistic experience with lexical tones and musical training should have similar and comparable effects on tone perception performance. Our findings that merging Cantonese speakers performed as well as the standard Cantonese speakers on the pure tone task is expected by her hypothesis. Nevertheless, the findings in the present study also raise a number of questions about this hypothesis. First, if lexical tones are treated by the brain in exactly the same way as musical tones, why would there be different performance in the two domains by the same native speakers? Specifically, the standard Cantonese listeners performed better than the merging Cantonese listeners in lexical tone perception but performed similarly in pure tone perception. Such discrepancy between lexical and musical tones is not predicted by the same neural circuitry hypothesis. One possible reason may be related to the higher-level cognitive processes involved in lexical processing but not in musical processing. Presumably, the standard Cantonese listeners could better make use of top-down lexical information than the merging listeners in the lexical tone task. More studies targeting processing at different levels for native tone speakers are needed to elucidate this possibility.

Second, the existence of musicians who are merging lexical tones is also not expected by Deutsch's hypothesis. Although the effects of lexical tone experience and musical training are not incremental in her hypothesis, it is reasonable to expect that their effects should at least be similar and comparable in the same direction. It is particularly worth pointing out that among the 10 merging Cantonese musicians in this study, three were at professional performance level. Two of them majored in Music at the Chinese University of Hong Kong. One majored in piano performance; the other held a performance diploma in vocal singing in addition to having studied the Guzheng (a Chinese string instrument) for more than 10 years, and the piano for more than 6 years. The third merging musician also held a performance diploma in piano. Nevertheless, such advanced musical training still could not prevent them from merging lexical tones. Or conversely, the merging of lexical tones did not hamper their musical abilities at all. These intriguing cases demonstrate that linguistic tones and musical tones were treated separately by them. In addition to the existence of merging musicians, the finding of no association between the ability of absolute pitch and the ability to use F0 height to identify Taiwanese tones by Taiwanese musicians by Lee *et al.* (2011) does raise some questions to the same circuitry hypothesis. Clearly, more studies specifically focusing on musicians who speak a tone language natively are needed to further explore this interesting area.

One possible interpretation of the perceptual differences between standard Cantonese and merging Cantonese subjects on lexical tones is that they speak different "dialects" of Cantonese, analogous to the differences between British and American English. Speakers of one dialect (merging Cantonese) would do less well in making judgments on the other dialect (standard Cantonese) than on their own dialect. There are indeed dialects of Cantonese, for example, Hong Kong Cantonese versus Guangzhou Cantonese. In addition to the high-level tone T1 [55], there is a high-falling tone [53] in Guangzhou Cantonese, but this variant is already assimilated to or is used in free variation with the high-level realization [55] in Hong Kong Cantonese (So, 1996; Bauer and Benedict, 1997). Nevertheless, the tonal variations between standard Cantonese and merging Cantonese subjects are not the same as those of dialectal differences. First of all, these speakers lived and grew up in the same place with comparable linguistic backgrounds (see Sec. II A). They still have the same phonological tone categories although there are differences in phonetic realizations. There is also large individual variation among the merging Cantonese speakers in terms of tone merge (e.g., different tone pairs and degrees of merging), which indicate that merging Cantonese is not, as it were, an "established dialect." Merging Cantonese can at best be viewed as subvarieties of standard Cantonese (Bauer *et al.*, 2003). Therefore dialectal differences cannot explain the weaker performance of merging Cantonese subjects on the lexical tones.

There is much recent independent evidence in the literature supporting the separation of musical and linguistic domains from different perspectives. While our results show no musical facilitation on lexical tone perception by native Cantonese speakers, there are studies showing no transfer effect from lexical tone experience to music perception in behavioral tasks. Nan *et al.* (2010) and Jiang *et al.* (2010) found that Mandarin speakers may experience musical pitch disorder (congenital amusia) in a way similar to nontone language speakers despite their early exposure to speech-related pitch contrasts. For example, the prevalence of congenital amusia in Mandarin speakers in Nan *et al.*'s study (3.4%) is very close to the percentage reported in Western countries (4%). Mandarin amusic subjects in both studies have similar musical impairment as Western amusic subjects. They also had problems in identifying both speech and non-linguistic analogues for Mandarin intonation. Thus tone language experience provides little compensation for the pitch disorder, although more recently Wong *et al.* (2012) found that Cantonese amusic participants showed enhanced pitch abilities relative to their Canadian amusic counterparts. Differences in the target languages (Mandarin vs Cantonese) and in the experimental tasks (identification and discrimination of lexical tones and melodic contours vs (in)congruity judgments of musical melodies) may explain the different findings in the preceding studies on amusic tone language speakers.

Despite studies showing that knowledge of a tone language can enhance pitch perception (e.g., Pfordresher and Brown, 2009), fluency in a tone language may interfere with, rather than facilitate, the perception of falling pitches in a non-speech context. Both Bent *et al.* (2006) and Peretz *et al.* (2011) found that native speakers of a tone language were impaired in the discrimination of falling pitches in tone sequences in non-speech context as compared to speakers of a nontone language. They argued that tone language listeners may have relied on their speech-specific processes as compensatory strategy when pitch differences are difficult to hear. In any case, their results indicate that, counter-intuitively, tone language experience does not necessarily enhance pitch perception, demonstrating the separation between linguistic and non-linguistic (musical) domains.

It seems possible that musical training can facilitate linguistic tone perception only if musical training starts before the linguistic use of tones as in the case of nontone musicians. Explicit musical training has increased their sensitivity to detect subtle pitch differences in linguistic tones more accurately (Wong *et al.*, 2007; Bidelman *et al.*, 2010). However, if the linguistic use of tones has started already, as in the case of the merging Cantonese subjects, even advanced musical training cannot enhance their perception of linguistic tones at all. Our findings imply that the linguistic use of tones is more fundamental and more robust than musical tones. If the linguistic domain is not activated, then musical training may enhance lexical tone perception but not vice versa.

The suggestion of the more fundamental role of the linguistic domain is supported by recent studies. Tillmann *et al.* (2011) found that in the presence of severe musical pitch deficit (congenital amusia), pitch information was slightly better perceived in speech materials (syllables) than on musical analogs. In addition, Cooper and Wang (2012) showed that either musical experience (for English listeners) or a tone language background (Thai listeners) led to significant improvement in non-native tone word learning in Cantonese as compared with those without musical training or tone language experience. However, musical background did not provide any additional advantage for Thai musicians over Thai non-musicians in learning non-native Cantonese words. Their findings suggest a differential in relevance of musicality depending on linguistic background.

Perhaps the best way to incorporate the seemingly discrepant results in the literature on the relationship between music and speech is to view the two domains as separate but overlapping at different levels. One possibility is that there can be domain-general perceptual processes but domain-specific structural representations between music and speech (Patel, 2003, 2008). Although this idea was proposed based on syntactic processes and representations, Wong *et al.* (2012) suggested that the same proposal can also be used in tone perception. A second possibility is that music and speech share lower-level processes while maintain separate higher-level processes. Nan and Friederici (2012) compared pitch processing in music and tone language in Mandarin-speaking musicians with fMRI. They found both shared neural circuits for pitch processing across domains and domain-specific modulation of the perceptual processes. Similarly, Bidelman *et al.* (2010, 2011) showed that there can be shared lower-level circuitry in the brain stem (pre-attentive stage) but different higher-level processing (cognitive stages) between music and speech. They argued that perceptual benefits across domains only exist when the information is behaviorally relevant to the listeners. Möttönen *et al.* (2005) found different neural activities in the same brain area (left STSp) when the subjects listened to identical sounds (sine wave speech) as speech or non-speech. All these results suggest that more brain imaging studies can reveal the complex interactions and mechanisms between music and speech and the mental architecture of tone perception in general.

To conclude, the current study investigated the effects of musical training on Cantonese tone perception by native and nontone speakers. Musical training enhanced lexical tone perception of nontone speakers while it had little effect on both standard and merging Cantonese speakers. The findings suggest that the linguistic and musical mechanisms belong to separate but overlapping domains.

ACKNOWLEDGMENTS

The authors would like to thank Alan Yu for collecting nontone musician data for us. They also thank Peggy Wong for her contribution in the early stages of the project, and Donald White for editing the manuscript. Special thanks also go to the reviewers for very helpful comments and suggestions in previous versions of the manuscript.

Bauer, R., Cheung, K. H., and Cheung, P. M. (**2003**). "Variation and merger of the rising tones in Hong Kong Cantonese," Lang. Var. Change **15**, 211–225.

Bauer, R. S., and Benedict, P. K. (**1997**). *Modern Cantonese Phonology* (Mouton de Gruyter, Berlin), Chap. 2.

Bent, T., Bradlow, A. R., and Wright, B. A. (**2006**). "The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds," J. Exp. Psychol. Hum. Percept. Perform. **32**, 97–103.

Bidelman, G. M., Gandour, J., and Krishnan, A. (**2010**). "Cross-domain effects of music and language experience on the representation of pitch in the human auditory brain stem," J. Cognit. Neurosci. **23**, 425–434.

Bidelman, G. M., Gandour, J., and Krishnan, A. (**2011**). "Musicians and tone-language speakers share enhanced brain stem encoding but not perceptual benefits for musical pitch," Brain Cogn. **77**, 1–10.

Burnham, D., and Mattock, K. (**2007**). "The perception of tones and phones," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O. S. Bohn and M. J. Munro (John Benjamins, Amsterdam), pp. 259–280.

Chandrasekaran, B., Krishnan, A., and Gandour, J. (**2009**). "Relative influence of musical and linguistic experience on early cortical processing of pitch contours," Brain Lang. **108**, 1–9.

Cooper, A., and Wang, Y. (**2012**). "The influence of linguistic and musical experience on Cantonese word learning," J. Acoust. Soc. Am. **131**, 4756–4769.

Deutsch, D. (**2002**). "The puzzle of absolute pitch," Curr. Dir. Psychol. Sci. **11**, 200–204.

Deutsch, D., Dooley, K., Henthorn, T., and Head, B. (**2009**). "Absolute pitch among students in an American music conservatory: Association with tone language fluency," J. Acoust. Soc. Am. **125**, 2398–2403.

Deutsch, D., Henthorn, T., and Dolson, M. (**2004**). "Absolute pitch, speech, and tone language: Some experiments and a proposed framework," Music Percept. **21**, 339–356.

Deutsch, D., Henthorn, T., Marvin, E., and Xu, H. (**2006**). "Absolute pitch among American and Chinese conservatory students: Prevalence, differences, and evidence for a speech-related critical period," J. Acoust. Soc. Am. **119**, 719–722.

Ettlinger, M., Margulis, E. H., and Wong, P. C. M. (**2011**). "Implicit memory in music and language," Front. Psychol. **2**, 211.

Gandour, J. (**1981**). "Perceptual dimensions of tone: Evidence from Cantonese," J. Chin. Linguist. **9**, 20–36.

Gottfried, T. L. (**2007**). "Music and language learning: Effect of musical training on learning L2 speech contrasts," in *Language Experience in Second Language Speech Learning*, edited by O. S. Bohn and M. Munro (John Benjamins, Philadelphia), pp. 221–237.

Hallé, P., 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.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., and Yang, Y. (**2010**). "Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese," Neuropsychologia **48**, 2630–2639.

Khouw, E., and Ciocca, V. (**2007**). "Perceptual correlates of Cantonese tones," J. Phonetics **35**, 104–117.

Lee, C. Y., and Hung, T. H. (**2008**). "Identification of Mandarin tones by English-speaking musicians and nonmusicians," J. Acoust. Soc. Am. **124**, 3235–3248.

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.

Mok, P., and Wong, P. W. Y. (**2010a**). "Perception of the merging tones in Hong Kong Cantonese: Preliminary data on monosyllables," *Proceedings of Speech Prosody 2010*, Chicago, IL, pp. 100916:1–4.

Mok, P., and Wong, P. W. Y. (**2010b**). "Production of the merging tones in Hong Kong Cantonese: Preliminary data on monosyllables," *Proceedings of Speech Prosody 2010*, Chicago, IL, pp. 100986:1–4.

Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., and Sams, M. (**2005**). "Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus," Neuroimage **30**, 563–569.

Nan, Y., and Friederici, A. D. (**2012**). "Differential roles of right temporal cortex and broca's area in pitch processing: Evidence from music and Mandarin," Hum. Brain Mapp, in press (2012).

Nan, Y., Sun, Y., and Peretz, I. (**2010**). "Congenital amusia in speakers of a tone language: Association with lexical tone agnosia," Brain **133**, 2635–2642.

Patel, A. D. (**2003**). "Language, music, syntax and the brain," Nat. Neurosci. **6**, 674–681.

Patel, A. D. (**2008**). *Music, Language and the Brain* (Oxford University Press, Oxford), Chap. 2.

Peretz, I., Nguyen, S., and Cummings, S. (**2011**). "Tone language fluency impairs pitch discrimination," Front. Psychol. **2**, 145.

Pfordresher, P. Q., and Brown, S. (**2009**). "Enhanced production and perception of musical pitch in tone language speakers," Attention, Percept. Psychophys. **71**, 1385–1398.

Schellenberg, E. G., and Peretz, I. (**2008**). "Music, language and cognition: Unresolved issues," Trends Cognit. Sci. **12**, 45–46.

So, L. K. H. (**1996**). "Tonal changes in Hong Kong Cantonese," Curr. Issues Lang. Soc. **3**, 186–189.

Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., and Peretz, I. (**2011**). "Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages," Front. Psychol. **2**, 120.

Vance, T. J. (**1977**). "Tonal distinctions in Cantonese," Phonetica **34**, 93–107.

Varley, R., and So, L. K. H. (**1995**). "Age effects in tonal comprehension in Cantonese," J. Chin. Linguist. **23**, 76–97.

Wang, Y., Behne, D. M., Jongman, A., and Sereno, J. A. (**2004**). "The role of linguistic experience in the hemispheric processing of lexical tone," Appl. Psycholinguist. **25**, 449–466.

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

Wayland, R., Herrera, E., and Kaan, E. (**2010**). "Effects of musical experience and training on pitch contour perception," J. Phonetics **38**, 654–662.

Wong, P. C. M., Ciocca, V., Chan, A. H. D., Ha, L. Y. Y., Tan, L. H., and Peretz, I. (**2012**). "Effects of culture on musical pitch perception," PLoS ONE **7**, e33424.

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (**2007**). "Musical experience shapes human brain stem encoding of linguistic pitch patterns," Nat. Neurosci. **10**, 420–422.