I. INTRODUCTION

The challenges that face adult second language (L2) learners have been well documented, and a variety of experiential factors have been found to influence how challenging certain phonemic contrasts are to acquire (e.g., Best and Tyler, 2007; Flege, 1995; Werker and Tees, 2002). Linguistic background is considered a major factor in determining learner performance, including the amount of L1 (Flege and MacKay, 2004) and L2 use (e.g., Flege, Bohn, and Jang, 1997) and the age of learning (e.g., Flege, Munro, and MacKay, 1995), as well as the nature of the relationship between the native language (L1) phonetic inventory and the L2 inventory (e.g., Best, McRoberts, and Goodell, 2001; Flege, 1995). Native language background can potentially facilitate or inhibit L2 perception (e.g., Aoyama, Flege, Guion, Akahane-Yamada, and Yamada, 2004; Best et al., 2001). The inhibitory versus facilitative influence of native categories can be viewed in the context of L2 segmental theories such as the Perceptual Assimilation Model (e.g., Best, 1995; Best and Tyler, 2007) and the Speech Learning Model (e.g., Flege, 1995, 2007), which posit that some contrasts will prove more challenging than others, depending on the degree of perceived acoustic or gestural similarity between the L2 contrast and the L1 category. Moreover, a variety of non-linguistic factors, including neurological maturation, learner motivation and attitude are also purported to impact learning (e.g., Gardner, Lalone, and Moorcroft, 1985; Lenneberg, 1967). One particular extralinguistic factor, musical experience, has also been shown to contribute to L2 speech learning, as enhanced auditory acuity developed as a product of long-term musical training may aid listeners in discerning difficult non-native distinctions (e.g., Slevc and Miyake, 2006).

A. Effects of language background

Linguistic experience influences not only non-native segmental but also suprasegmental perception (Francis, Ciocca, Ma, and Fenn, 2008; Gandour, 1983; Gandour, Wong, Lowe, Dzemidzic, Satthamnuwong, Tong, and Li, 2002; Gottfried and Suiter, 1997; Wang, Jongman, and Sereno, 2006; Wayland and Guion, 2004). Many of the world’s languages, such as Mandarin, Cantonese, and Thai, use different pitch patterns (tones) systematically to differentiate words, and these tones are often the only element that can distinguish words from each other (Yip, 2002). Non-native tones can be a challenging element to perceive, particularly for listeners whose native languages are non-tonal (e.g., English) and who are consequently less accustomed to attending to lexically meaningful pitch information (Lee, Vakoch, and Wurm, 1996; Wayland and Guion, 2004; Wayland and Li, 2008). However, studies have demonstrated that linguistic experience can mediate how certain tones are perceived. Gandour (1983) reported that perceptual discrepancies between language groups could be attributed to the relative weighting of two perceptual dimensions: fundamental frequency (F0) height and direction of change. Non-tone language listeners (English) attached less importance to the “direction of change” dimension and gave greater weight to the “height” dimension; whereas, “direction of change”
perceiving non-native tones (Francis et al., 2008; Wang, Spence, Jongman, and Sereno, 1999; Wayland and Guion, 2004). For example, Wayland and Guion (2004) trained native Chinese and English groups on the mid and low tonal contrast in Thai, reporting that the Chinese listeners had higher discrimination accuracy even before training, and that they were the only group to significantly improve from pre- to post-test. They suggested that native language experience with detecting word-level F0 changes could have positively transferred to aid non-native tone discrimination. Neurophysiological findings support this notion, as pitch-tracking accuracy in the brainstem while listening to Mandarin and Thai tones was more accurate for Mandarin and Thai listeners for both linguistic stimuli types as compared to English listeners (Krishnan, Gandour, and Bidelman, 2010).

Despite the challenges facing most non-native listeners, particularly non-tone language listeners, previous research have demonstrated that adults are capable of improving their ability to distinguish non-native tones, even after only a brief period of laboratory training (e.g., Francis et al., 2008; Wang, Spence, Jongman, and Sereno, 1999; Wayland and Li, 2008). For example, Wang et al. (1999, 2003), utilizing high-variability procedures, trained American English listeners to identify the four Mandarin tones, which resulted in a significant improvement in accuracy after training, with the improvement generalizing to new stimuli, novel voices, and to the production domain.

However, other studies suggest that native tone language experience alone does not necessarily predict greater success over those with non-tone language backgrounds in perceiving non-native tones (Francis et al., 2008; Wang, 2006). For example, Francis et al. (2008) tested native Mandarin and English listeners on Cantonese tone perception before and after perceptual training. Unlike Wayland and Guion (2004), no significant difference in identification accuracy was found between Mandarin and English listeners on both the pre- and post-tests, suggesting that a tone language background is not necessarily advantageous for non-native tone perception. However, group differences became evident in their particular tonal confusions and tonal accuracy patterns. For instance, consistent with the notion that native category representations can have a significant influence on non-native perception (e.g., Flege, 1995), Mandarin listeners’ performance was best on the three Cantonese tones that have similar counterparts in Mandarin. Their findings suggest that the composition of the native tonal or intonational inventories and their relationship to incoming non-native tones, such as L1 (Mandarin) tones assimilating with non-native (Cantonese) tones, can better explain differences in cross-linguistic tone perception than simply the presence or absence of native lexical tones.

While most of these training studies have focused on improving listeners’ ability to identify or discriminate specific contrasts, research has also examined whether the ability to discern non-native acoustic distinctions can be applied to broader linguistic contexts such as word learning. Two studies in particular have examined how listeners learn to use non-native contrasts in lexical identification tasks (Curtin, Goad, and Pater, 1998; Wong and Perrachione, 2007), reporting that participants were capable of being trained to use specific phonemic distinctions to distinguish foreign vocabulary items. Specifically, Curtin et al. (1998), comparing English and French listeners’ learning of Thai words differing in stop consonant aspiration and voicing contrasts, concluded that listeners initially form lexical representations utilizing non-native features contrastive in their L1 (in this case voicing), before lexicalizing features that are not present or phonemically relevant in their L1 phonology (aspiration). Of particular relevance to the study of tone word learning, Wong and Perrachione (2007) examined the ability of English listeners to learn English pseudowords distinguished by three Mandarin lexical tones. They reported that participants were capable of using non-native tonal contrasts to distinguish word meaning, and that tonal awareness, as determined by a pre-training pitch pattern identification task, was an important factor in learning the meaning of words distinguished by pitch patterns.

B. Effects of musical experience

In addition to linguistic experience, musical background has also been shown to be particularly influential on suprasegmental prosodic perception, due in large part to the overlap between prosody and music in the relevant acoustic features, such as fundamental frequency and duration. Numerous studies have found an advantage of musicianship for non-tone language participants identifying and discriminating non-native lexical tones (Alexander, Wong, and Bradlow, 2005; Delogu, Lampis, and Belardinelli, 2006, 2010; Gottfried, 2007; Lee and Hung, 2008). Delogu et al. (2010), for example, found that musical expertise predicted better non-native tone identification in Italian adults and children; although, it did not appear to substantively impact non-native segmental identification. Similarly, English musicians displayed significantly better accuracy rates than non-musicians when identifying Mandarin lexical tones on intact, silent-center and onset-only syllables (Lee and Hung, 2008). There is also evidence pointing to neurophysiological changes as a result of long-term exposure to musical pitch and its effect on linguistic pitch processing. Wong, Skoe,
Russo, Dees, and Kraus (2007) reported English musicians as having a more robust encoding of linguistic pitch in the auditory brainstem when listening to Mandarin lexical tones, suggesting that long-term pitch exposure may alter fundamental sensory circuitry. The effect of musicianship was also found in tone word learning, as Wong and Perrachione (2007) reported that musicians were more successful than non-musicians at learning to use pitch for lexical identification. Thus, there appears to be robust evidence indicating that musical experience enhances auditory acuity, which can aid in the perception of difficult non-native prosodic contrasts, such as lexical tone.

C. Interaction of linguistic and musical experience

While a substantial body of work has examined the influence of linguistic and musical experience on lexical tone perception separately (e.g., Delogu et al., 2010; Wayland and Guion, 2004), only a few studies have investigated the interaction of these factors. Recent research compared the effects of musical training and a tone language background on the neural processing of pitch, as reflected by pitch-tracking accuracy in the auditory brainstem (Bidelman, Gandour, and Krishnan, 2011b). Native Mandarin, English musician and non-musician groups were presented with homologues of a Mandarin lexical tone (rising tone 2) and a musical interval (major third). Results revealed that both English musicians and native Mandarin listeners showed higher pitch-tracking accuracy and pitch strength relative to English non-musicians, indicating that both musical and linguistic experience similarly shaped pitch processing. Correspondingly, Chandrasekaran, Krishnan, and Gandour (2009) found that Mandarin tone homologues presented in a non-speech context produced greater mismatch-negativity responses in English musicians and native Mandarin listeners as compared to English non-musicians. These findings point to domain-general mechanisms for pitch processing which can facilitate both linguistic and non-linguistic pitch processing.

D. The current study

Although previous studies have reported neurophysiological or behavioral enhancements for musicians and tone language speakers processing non-native linguistic tonal information (e.g., Chandrasekaran et al., 2009; Wong et al., 2007), the contributions of linguistic and musical experience are by no means straightforward, modulated by context and attention demands (e.g., Bidelman, Gandour, and Krishnan, 2011a; Tervaniemi, Kruck, De Baene, Schröger, Alter and, Friederici, 2009). Moreover, while previous research has examined the interaction of these factors, their focus has been on comparing tone language speakers perceiving native tone stimuli against non-native musicians and non-musicians (e.g., Bidelman et al., 2011b; Chandrasekaran et al., 2009). In order to investigate the interactive effects of musical and linguistic experience on non-native tone perception and word learning, the present study included groups of non-native listeners varying in both their L1 (tone versus non-tone language) and musical background. Examining their interaction may provide insight into how the cognitive mechanisms employed in processing non-native contrasts are shaped by different experiences. Furthermore, the majority of previous research has concentrated on the influence of experience on the perception and learning of individual speech contrasts (e.g., Bidelman et al., 2011b; Francis et al., 2008; Wang et al., 2003), and only a handful of studies have investigated the factors that affect listeners’ abilities to use these contrasts in a higher linguistic context, such as in words or phrases (Curtin et al., 1998; Wong and Perrachione, 2007). As second language learning requires listeners to be able not only to distinguish non-native contrasts but also crucially to utilize them in larger communicative contexts, research examining the relationship between sound-to-word learning and its influencing factors is important to our understanding of second language acquisition.

Thus, the present research focuses on the interaction between linguistic and musical experience in non-native (Cantonese) tone word learning by recruiting both native tone (Thai) and non-tone (English) language listeners with and without musical experience (i.e., Thai and English listeners, subdivided into musician and non-musician groups). The participants were asked to complete a word identification training program, learning to identify word meaning distinguished by Cantonese tones. They also completed a Cantonese tone identification task before and after training, as well as a musical aptitude task to establish their level of tonal awareness and auditory aptitude.

The Cantonese tonal inventory is comprised of six contrastive lexical tones, including three static (high-level, mid-level, low-level) and three contour tones (high-rising, low-rising, low-falling; Bauer and Benedict, 1997), perceptually distinguishable based on relative pitch (Vance, 1976). In contrast, Standard Thai has five lexical tones, which include three level (high, mid, low) and two contour tones (rising, falling; Abramson, 1962).

The primary goal of this study was to differentiate the relative influences of native language background and musical experience, examining whether one of these factors facilitated the acquisition of new lexical items in Cantonese to any greater degree than the other at the initial stage of non-native tone word learning. No difference found between either experiential factor (e.g., English musicians and Thai non-musicians) might provide further support for an overlap in the cognitive mechanisms utilized for both language and music. On the other hand, Thai non-musicians making larger gains in tone word identification accuracy than English musicians would suggest that the relevant processes necessary for feature-to-word mapping are more domain-specific.

Specifically, the present research predicted that the relative influence of linguistic and musical factors would produce a hierarchy of word learning success. First, Thai musicians were expected to have the highest success in learning the Cantonese words, resulting from an additive effect of musical and linguistic tone experience. Next, the performance of the Thai non-musicians and English musicians would be particularly interesting to compare, in that larger gains in one group would point to the relative weight of linguistic and musical experience in affecting word learning. We expected Thai non-musicians to make larger gains during training than the
English musicians and non-musicians. While English musicians have demonstrated superior pitch acuity (e.g., Alexander et al., 2005) and word learning proficiency (Wong and Perrachione, 2007) over English non-musicians, the Thais might have an advantage given their previous experience using tone in a lexically significant manner. Finally, in line with Wong and Perrachione (2007), the English musicians were predicted to be more successful during word training than their non-musician counterparts. The English musicians’ enhanced auditory perception should allow them to discern and retain the tonal contrasts with greater ease, enabling them to focus their attention on learning the appropriate meanings.

Concerning the tone identification task, given that previous findings have shown that linguistic experience can facilitate (e.g., Wayland and Guion, 2004) or inhibit tone perception (e.g., Francis et al., 2008; Wang, 2006), it was not certain whether Thais would have greater performance accuracy on Cantonese tone identification than the English group. However, given the results from Francis et al. (2008) on Cantonese tones specifically, where English listeners had similar accuracy rates as Mandarin listeners, we speculated that we might not find a significant difference in accuracy between the Thai and English groups. In terms of musical experience, based on findings that musicians were more accurate at identifying non-native tones (Alexander et al., 2005; Gottfried, 2007), musically experienced participants (both English and Thai) were expected to have higher levels of accuracy than their respective non-musician counterparts. Furthermore, considering previous research reporting that tonal awareness was a significant factor in tone word learning (Wong and Perrachione, 2007), it was also predicted that the scores on the tone identification task would be a significant predictor of word learning success.

II. METHODS

A. Participants

A total of 54 adults were included in the present study, including 28 native Thai speakers and 26 native English speakers. Participants were all college students, with self-reported normal hearing and cognitive abilities, and with no previous knowledge of Cantonese or any other lexical tone language aside from their native language. Each language group was comprised of two sub-groups: musician and non-musician. Musicians were defined as individuals who had undergone at least seven years of continuous Western instrumental music training and had a current ability to play an instrument, whereas non-musicians had no musical training within the last five years and less than two years of musical experience prior to that (Wayland, Herrera, and Kaan, 2010; Wong et al., 2007). Classical Thai music is traditionally learned from oral instruction or rote method (Morton, 1976). This style of learning may demand and develop a greater attentiveness to auditory information, as compared to students who learn from reading notation. Therefore, only Thai musicians who received their musical training on Western instruments (e.g., piano, violin) were permitted to participate in order to avoid discrepancies in training styles and musical scale systems between the Thai and English musician groups.

Specifically, the Thai participants were students from Chulalongkorn University and Silpakorn University in Bangkok, Thailand. All were native speakers of the Bangkok dialect (Standard Thai), with English as their primary foreign language. Fourteen were non-musicians (7 male, 7 female; mean age: 22 years), with an average of 0.21 years of musical experience (SD = 0.56). Fourteen participants fit the criteria of musician (6 male, 8 female; mean age: 21 years), with their amount of musical experience ranging from 7 to 18 years (M = 10.14 years, SD = 3.86). The native English participants were students from Simon Fraser University and the University of British Columbia in Vancouver, Canada. Fourteen participants were considered non-musicians (5 male, 9 female; mean age: 24), possessing an average of 0.69 years of musical experience (SD = 0.85). Twelve participants fell under the above-mentioned musicianship criteria (4 male, 8 female; mean age: 21), with the amount of musical experience ranging from 9 to 17 years (M = 11.91 years, SD = 2.68). A summary of the musicians’ characteristics is provided in Table I. All participants were paid for their participation in this study.

### Table I. Instrumental bios, cumulative years of musical experience, and the age of onset of musical training for the musician groups.

<table>
<thead>
<tr>
<th>Instrument(s)</th>
<th>Total duration of musical experience (yr)</th>
<th>Age of onset (yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English musicians</td>
<td>Thai musicians</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>Average</td>
</tr>
<tr>
<td></td>
<td>11 (SD: 3)</td>
<td>9 (SD: 4)</td>
</tr>
</tbody>
</table>

The native English and Thai native groups were matched on age (M = 11.91 years, SD = 2.68). A summary of the musicians’ characteristics is provided in Table I. All participants were paid for their participation in this study.

<table>
<thead>
<tr>
<th>Instrument(s)</th>
<th>Total duration of musical experience (yr)</th>
<th>Age of onset (yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English musicians</td>
<td>Thai musicians</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>Average</td>
</tr>
<tr>
<td></td>
<td>11 (SD: 3)</td>
<td>9 (SD: 4)</td>
</tr>
</tbody>
</table>
In order to ensure (1) a distinct separation between the musician and non-musician groups and (2) the comparability of the Thai and English groups in terms of their years of musical experience, a one-way analysis of variance (ANOVA) on years of musical experience with participant group as the independent variable was conducted. The results yielded a significant main effect of group [F(3,53) = 88.358, p < 0.0001]. Post hoc analyses (Bonferroni) revealed that, as expected, musicians had significantly more years of musical training than non-musicians (p < 0.0001). Crucially, no significant difference in years of musical experience between the two musician groups (p = 1.00) or the two non-musician groups (p = 0.97) was found. Additionally, the comparability of the cross-domain (linguistic and musical background) comparisons was evaluated. Ciocca and Lui (2003) reported that Cantonese-speaking children acquire their native tone system, achieving adult levels of performance, by approximately 10 years of age, suggesting that continued exposure to native tones would not markedly improve their tone perception abilities. If native tone language speakers take approximately 10 years to acquire their native tone system, then this is comparable to the years of musical experience of this study’s participants. Furthermore, given that English musicians and tone language speakers have demonstrated comparable levels of enhanced pitch encoding relative to English non-musicians (Bidelman et al., 2011b), it seemed reasonable to make the intended group comparisons of language background and musicality.

B. Stimuli

Stimuli were recorded in a sound-attenuated booth in the Language and Brain Lab at Simon Fraser University using a Shure KSM109 microphone. The recordings were made using Sound Forge 3.0 (Sonic Foundry, 1995) and the Presonus Firebox audio interface at a 44.1 kHz sampling rate.

1. Pre-/post-training tone identification

Two native Hong Kong Cantonese speakers (1 male, 1 female) produced a set of stimuli for the pre- and post-training tone identification task. Five Cantonese CV monosyllables (waj, low, si, pej, fu) containing common Thai/English phonemes were selected and produced with five Cantonese tones, for a total of 25 stimuli. The five Cantonese tones selected for use in this study were high-level (55), high-rising (25), low-falling (21), low-rising (23), and low-level (22) (Chao, 1947). The mid-level (33) tone was not included to reduce the difficulty level, as it would likely be more easily confused with high-level, low-level, or low-rising tones, in the absence of any contextual cues (Francis et al., 2008).

2. Tone word training

For use in the training program, four native Cantonese speakers (2 male, 2 female) not used in the tone identification task produced three Cantonese CV monosyllables (k`aj, tsou, wu) with five Cantonese tones. The syllables were comprised of a Cantonese initial consonant and a vowel common to Cantonese and the participants’ native languages (Thai and English), as studies have found that word learning can be inhibited when learners are faced with unfamiliar phonotactics (e.g., Ellis and Beaton, 1993a). In order to reduce lexical competition, these syllables were selected as they do not contain semantic content in either Thai or English.

Each word was assigned a distinct meaning, as represented by a picture. The imageability of a concept has been shown to facilitate retrieval cues, and depicting word meanings as images might aid learning to a greater degree than providing their translated cues (Chun and Plass, 1996). Specifically, concrete nouns were chosen as the meanings for the pseudowords, as they have been found to be easier to remember (Ellis and Beaton, 1993b). These pictures were selected from a set of 260 standardized pictures, controlled for visual complexity and cultural familiarity (Snodgrass and Vanderwart, 1980). Three categories of pictures were included for the word learning program (animals, human body parts, household items).

C. Procedure

Participants completed all of the tasks on PC computers in a quiet room. Aural stimuli were played through headphones at a comfortable listening volume. Task instructions and feedback information were provided in English and Thai for the respective participant groups. The test and training procedures were administered in the following sequence: musical aptitude task, pre-training tone identification task, tone word training and session tests, and post-training tone identification task.

1. Musical aptitude task

Prior to the tone identification task and word training program, participants first completed a musical aptitude task. The Advanced Measures of Music Audiation (AMMA; Gordon, 1989) was administered, as it was designed specifically for high-school students and adults such that no prior music training was required. The task involved listening to 30 pairs of piano melodies and indicating whether they were identical or different on a paper answer sheet. If the latter, they specified whether it was a tonal or rhythmic variation. Participants were first provided three practice trials demonstrating the three possible melodic pairs (same, tonal difference, rhythmic difference). The task took approximately 20 min to complete.

2. Pre-/post-training tone identification task

All groups completed a tone identification task before and after the training program. A familiarization portion was employed prior to the tone identification test to allow participants to become familiar with the five Cantonese tones and learn how to identify them. This task utilized one syllable (ji) recorded by the female speaker from the pre-/post-training tone identification task. They first heard each Cantonese tone pronounced in isolation and viewed an associated tone diagram (a visual representation of the contour/level tone). The participants were then asked to respond after each stimulus, identifying the tone they heard by pressing the number corresponding to the appropriate tone diagram. Feedback on response accuracy and the correct answer were provided.
Fifteen trials comprised this task (5 tones × 1 syllable × 1 speaker × 3 repetitions), which was considered sufficient time to be familiarized with the task and tones (Wayland and Li, 2008).

Next, they completed the main tone identification task to determine their ability to distinguish the non-native tones. The format was identical to the familiarization task, only they did not receive feedback. They identified 100 stimuli (5 syllables × 5 tones × 2 speakers × 2 repetitions), presented with an inter-stimulus-interval of 3 s, by selecting the appropriate pitch pattern as presented on the screen.

3. Tone word identification training task

The participants completed seven 30-min training sessions on four days spread out over two weeks. The first three training days contained two training sessions per day, with a 15-min break between sessions, and the last day contained one training session. Training procedures were modeled after training provided in Curtin et al. (1998) and Wong and Perrachione (2007).

Each training session was comprised of five training blocks, followed by two review blocks and a session test. The format of each training block consisted of listening to four randomized repetitions (two repetitions each from a male and female speaker) of three words while viewing the visual representation of their meaning. This task was intended to simulate vocabulary learning, in that listeners would have to map a new phonological form onto a pictorially represented meaning. Individual training blocks contained no minimal pairs, using three different tones on three different syllables [e.g., kwaj (55), tsou (23), wu (22)]. Additionally, the words in each block had picture meanings that were as semantically distinct as possible, as studies have suggested that semantic clustering of new vocabulary items may actually be a detriment to learning (Finkbeiner and Nicol, 2003; Tamminen and Gaskell, 2008). Each block concluded with a small quiz on all three words learned in that block (Fig. 1). They heard a stimulus and were presented with the three pictures of the words they had just learned. They were asked to indicate the correct meaning for the word by selecting the appropriate picture. In total, each block consisted of 12 listening trials (2 speakers × 3 words × 2 repetitions) and 12 quiz trials (4 speakers × 3 words).

After completing the five training blocks, participants received two additional blocks reviewing the lexical items. Review 1 was comprised of all 15 words, blocked by syllable, produced by a female speaker from the training blocks (totaling 15 trials). Blocking for syllable enabled participants to hear minimal quintuplets in succession in order to draw their attention to the tonal distinctions. Participants heard a word and identified its meaning from a choice of six options, a minimal tonal quintuplet plus a foil image. For review 2, participants heard two randomized repetitions of all 15 words from two training block speakers (totaling 30 trials), now choosing the appropriate meaning from all 15 potential options on the screen. For the block quizzes and both reviews, there was no limit on response time. Participants were also provided with feedback, informing them whether or not their answer was correct, displaying the correct answer and re-playing the aural stimulus.

At the end of each session, participants completed a session test on all 15 words learned in the training program without feedback, following a similar format as review 2. Participants identified four randomized repetitions (1 production from each of the four speakers) of all 15 tone words (totaling 60 trials) with an inter-stimulus-interval of 10 s. The results from this test were used to determine the degree of improvement for each participant.

III. RESULTS

A. Tone word learning

Tone word identification performance was assessed from the session tests concluding each training session. To evaluate overall improvement over the course of training, mean percent correct was calculated for the first (session 1) and last (session 7) training sessions (Fig. 2). These scores were first submitted to a three-way mixed-design ANOVA with session (session 1 test, session 7 test) as repeated measures and language (English, Thai) and musicality (musician, non-musician) as between-subjects factors. A significant main effect of session was obtained [F(1,50) = 498.926, p < 0.0001], indicating that listeners were able to significantly improve their word identification accuracy scores after training (from 24% to 67%). The ANOVA also yielded significant session × language [F(1,50) = 4.631,

FIG. 1. Sample training block including the training trials and block quiz. Note that the IPA and tone numbers denote the auditory stimulus played for the listeners (they did not appear on the screen).
between TM (63%) and any of the other groups (significant improvement as a result of training). The independent variable confirmed that each group made a highly significant change. Differences found between EM and TNM (75%) had significantly higher accuracy than ENM (54%; Bonferroni) analysis revealed that by the last session, EM performed significantly better than TNM (46%) across tests (p = 0.008). No significant differences were found between TNM and ENM (51%, p = 1.00), or between TM and ENM (p = 0.179).

Finally, the main effect of Tone [F(4, 49) = 27.547, p < 0.0001] and tone × language interaction were also significant [F(4, 49) = 9.606, p < 0.0001]. No significant tone × musicality [F(4, 49) = 1.215, p = 0.306] or tone × language × musicality interactions were found [F(4, 49) = 0.947, p = 0.438]. To examine the language group differences for which tones were easier or more challenging, one-way ANOVAs for each language with tone as a repeated measure were conducted. Bonferroni-adjusted pairwise comparisons revealed that Thai listeners’ highest identification scores were for low-falling tones (64%), significantly greater than low- (43%) and high-rising (44%) tones (p < 0.001). Low- (59%) and high-level (56%) tones were also identified with significantly greater accuracy than rising tones overall (p < 0.051).

B. Pre-/post-training tone identification

1. Percent correct

The mean percent correct tone identification scores were submitted to a mixed four-way ANOVA, with test (pre-training, post-training) and tone (high-level, high-rising, low-falling, low-rising, low-level) as repeated measures, and language and musicality as between-subjects factors. For brevity, only the results involving the group factors (language, musicality) will be reported. Group comparisons are illustrated in Fig. 3.

First, as expected, a significant main effect of test was obtained [F(1, 49) = 34.290, p < 0.0001], indicating an overall improvement in tone identification accuracy from pre-training (53%) to post-training (66%) across groups. Significant main effects of language [F(1, 49) = 17.414, p < 0.0001] and musicality [F(1, 49) = 53.356, p < 0.0001] as well as a language × musicality interaction [F(1, 49) = 6.381, p = 0.015] were also found, although there was no significant interaction of test × musicality × language [F(1, 49) = 0.394, p = 0.533]. To address the language × musicality interaction with all the possible group comparisons, a one-way ANOVA with group (ENM, EM, TNM, TM) as the independent variable was performed, yielding a significant main effect [F(3, 49) = 24.707, p < 0.0001]. Post hoc (Bonferroni) analysis found that across tests, EM had significantly higher accuracy rates (79%) than all other groups (p < 0.0001). TM also performed significantly better (60%) than TNM (46%) across tests (p = 0.008). No significant differences were found between TNM and ENM (51%, p = 1.00), or between TM and ENM (p = 0.179).

Finally, the main effect of Tone [F(4, 49) = 27.547, p < 0.0001] and tone × language interaction were also significant [F(4, 49) = 9.606, p < 0.0001]. No significant tone × musicality [F(4, 49) = 1.215, p = 0.306] or tone × language × musicality interactions were found [F(4, 49) = 0.947, p = 0.438]. To examine the language group differences for which tones were easier or more challenging, one-way ANOVAs for each language with tone as a repeated measure were conducted. Bonferroni-adjusted pairwise comparisons revealed that Thai listeners’ highest identification scores were for low-falling tones (64%), significantly greater than low- (43%) and high-rising (44%) tones (p < 0.001). Low- (59%) and high-level (56%) tones were also identified with significantly greater accuracy than rising tones overall (p < 0.051).
English groups had the most success with identifying high-level tones across tests (82%, \( p < 0.0001 \)). English listeners also identified low-level (64%) tones significantly better than low-rising overall (56%, \( p = 0.028 \)).

2. Tone confusion patterns

To further address the language \( \times \) tone interaction, tonal confusion analyses were also conducted, in order to gain additional insight into how L1 experience influences tone perception (e.g., Gandour, 1983; Francis et al., 2008). Confusion matrices were constructed (Tables II and III), whereby for a given tone (heard stimulus), the proportions of misidentifications were tabulated by tone (response given). Four confusion patterns for each tone (e.g., 25 > 55, 25 > 21, 25 > 23, 25 > 22, where “ > ” denotes “confused as”) were each submitted to two-way ANOVAs with language as a between subjects factor and confusion pattern as repeated measures. For brevity, only significant language \( \times \) confusion pattern interaction results are reported here.

There were significant language \( \times \) confusion pattern interactions for the high-level \( [F(3,51) = 19.427, p < 0.0001] \), low-falling \( [F(3,51) = 19.325, p < 0.0001] \) and low-rising \( [F(3,51) = 7.174, p = 0.010] \) tones. Post hoc (Bonferroni) analyses showed that, for the high-level tone, Thai listeners tended to misidentify it most commonly as high-rising (\( p < 0.006 \)), whereas English listeners had no significant difference in confusion patterns (\( p > 0.060 \)). For the low-falling tone, Thai listeners confused it either as low-level or low-rising significantly more often than as high-rising and high-level (\( p < 0.0001 \)), with no significant difference between the proportion of low-level and low-rising misidentifications (\( p = 1.00 \)). English groups also most frequently confused it as the low-level tone (\( p < 0.0001 \)), but this misidentification was significantly more prevalent than for Thai listeners (\( p < 0.003 \)). Finally, for the low-rising tone, while Thai listeners exhibited a relatively even distribution of confusion patterns (\( p > 0.135 \)), for English listeners, high-rising and low-level were significantly more frequent confusion patterns than the other patterns (\( p < 0.0001 \)).

C. Pre-training tone identification and tone word learning

Previous work reported that pitch pattern identification accuracy significantly predicted tone word learning proficiency in English (Wong and Perrachione, 2007). In order to examine whether such findings would hold for Cantonese tone word learning, a linear regression model was employed on the data from all groups, with percent correct word identification scores from the last training session (session 7) as the dependent variable and pre-training tone identification scores as the predictor. Results indicate that pre-training tone identification scores were a significant predictor of tone word identification by the end of training \( [R^2 = 0.207, F(1,52) = 6.822, p = 0.012] \). However, further linear regression models for each language group (English, Thai; Fig. 4) revealed that pre-training scores significantly positively predicted word learning success for English listeners \( [R^2 = 0.587, F(1,24) = 34.086, p < 0.0001] \) but not for the Thai group \( [R^2 = 0.0, F(1,26) = 0.002, p = 0.968] \).

D. Musical aptitude

With regards to musical aptitude, the total raw scores from the AMMA (Gordon, 1989) for all four groups were transformed into percentile rankings based on the percentile rank norms provided by the AMMA. For example, a participant in the 90th percentile in terms of college student norms possesses an exceptional degree of musical aptitude as compared to someone in the 20th percentile. A one-way ANOVA with ranking scores as the dependent variable and group as the independent variable yielded a significant group difference \( [F(3,53) = 17.903, p < 0.0001] \). Post hoc (Bonferroni) analysis revealed that both musician groups (EM, TM) achieved significantly higher percentile rankings than both the non-musician groups (\( p < 0.0001 \)). The musician groups did not differ significantly from each other (EM = 76, TM = 74; \( p = 1.00 \)), nor did the non-musician groups (ENM = 50, TNM = 42; \( p = 0.97 \)).

In order to determine whether musical aptitude predicted word proficiency level, a linear regression model was constructed on the data from all groups with musical aptitude percentile ranking as the predictor and session 7 scores as the dependent variable. Across groups, aptitude scores were a significant predictor of word learning proficiency \( [R^2 = 0.118, F(1, 52) = 6.822, p = 0.012] \). However, regressions on each language group (Fig. 5) revealed that musical aptitude ranking was a significant predictor of success in the word learning task for the English group \( [R^2 = 0.452, F(1, 24) = 19.817, p < 0.0001] \), but not for the Thai group.
These results indicate that the level of overall attainment in tone word learning for English groups, but not the Thai groups, can in part be predicted by their level of musical aptitude.

Finally, as both musical aptitude and pre-training tone identification scores were significant predictors of performance on the word learning task for the English listeners, and musical aptitude and pre-training identification scores were significantly correlated with each other \( r = 0.620, p = 0.001 \), partial correlations were performed to examine the relative contributions of these factors as predictors of word learning proficiency. A partial correlation was first computed between musical aptitude scores and session 7 scores, controlling for pre-training tone identification. The result was a moderately significant positive correlation \( r = 0.391, p = 0.053 \). A partial correlation between pre-training tone identification and session 7 scores, when musical aptitude was controlled, was also significantly correlated \( r = 0.601, p = 0.001 \). These findings indicate that tone identification scores, and to a lesser extent musical aptitude scores, contributed significant unique variance to predicting session 7 scores for English listeners.

IV. DISCUSSION

A. Effects of linguistic experience

For the tone word identification task, TNM attained higher word learning proficiency than their English counterparts by the end of training, confirming our initial prediction of an effect of L1 background on tone word learning. The performance asymmetry between TNM and ENM provides general support for the notion that L1 background can have a substantive influence on non-native speech perception (e.g., Best, 1995; Best and Tyler, 2007; Flege, 1995, 2007). While these models focused their attention on the perception of individual segmental contrasts, these findings point to an influence of L1 background in a higher-level lexical prosodic context. Our results suggest that Thai listeners’ native language experience with using pitch to differentiate word meaning can beneficially transfer to aid the acquisition of non-native tone words. This notion is in line with Curtin et al. (1998), who posited that listeners first construct lexical representations utilizing contrasts that are lexical in their native language before features that are not used contrastively in their L1. Given this finding, this may explain why TNM, whose L1 does use pitch lexically, were able to achieve greater word learning success than their English-speaking counterparts by the final session. English listeners possess less experience with lexically significant pitch contrasts than tone language listeners. While English does utilize pitch to denote verb/noun shifts in some cases and even to make lexical distinctions such as lexical stress differences (e.g., INsight/inCITE), such instances are relatively uncommon (Cutler, 1986). The current results support the notion that tone language listeners’ pitch experience can be advantageous when acquiring foreign tone words, as their native experience with lexically significant pitch aided their ability make pitch-to-meaning associations more efficiently.

On the other hand, the tone identification results revealed no significant difference in performance accuracy between ENM and both Thai groups, across pre- and post-training tasks. These results do not confirm some of the previous findings that have suggested that having a tone language background is more advantageous than a non-tone language background for identifying non-native tones (e.g., Lee et al., 1996; Wayland and Guion, 2004). However, they are more consistent with studies such as Francis et al. (2008), who found no significant difference in accuracy on either the pre- or post-test between Mandarin and English listeners identifying Cantonese tones. The authors suggested that group differences in native categories and experience-dependent perceptual cue weightings were reflected in part by differences in tonal confusion patterns. In the present study, similar results were found in the tonal confusion analyses.

FIG. 4. Mean percent correct for session 7 against the pre-training tone identification data for English \( (R^2 = 0.587, \ p < 0.0001) \) and Thai \( (R^2 = 0.00, \ p = 0.968) \) groups.

FIG. 5. Mean percent correct for session 7 against the musical aptitude percentile rankings for English \( (R^2 = 0.452, \ p < 0.0001) \) and Thai \( (R^2 = 0.027, \ p = 0.415) \) groups.
(e.g., Tables II and III), which revealed certain group differences that could be attributed to L1 influence. For instance, English listeners misidentified the low-falling tone as low-level significantly more than any other tone; whereas, Thai listeners’ misidentifications of the low-falling tone were split between low-rising and low-level. These differences in confusion patterns may be derived from a discrepancy in the weighting of perceptual dimensions in the L1. For example, Gandour (1983) posited that English listeners attended to pitch height to a greater degree than direction, and tone language listeners, particularly Thais, were more attuned to contour rather than height. The present English listeners’ primary misidentification of low-falling as low-level might stem from their focusing on F0 height. Conversely, Thai listeners confusing low-falling as low-rising suggests they were attending to the changing contour direction (regardless of the actual direction of change) more than height. Francis et al. (2008) reported a similar result for Mandarin listeners, who also initially misidentified low-falling tones as low and high-rising tones. Additionally, across groups, high-rising and low-rising tones were the most challenging to master. However, this particular tone pair has been reported to be a difficult contrast for even native listeners to distinguish (Ciocca and Lui, 2003); thus, it is probable that the locus of this difficulty is at a lower acoustic level rather than a result of interference from linguistic experience.

**B. Effects of musical experience**

The overall attainment level of EM after training was significantly greater than ENM, consistent with the results from Wong and Perrachione (2007), indicating that long-term experience with musical pitch perception can be transferred to the linguistic domain. While musical experience has been shown to facilitate non-native tone identification (e.g., Alexander et al., 2005; Gottfried, 2007), the current results illustrate that it can also be beneficial at a higher linguistic level, namely tone word identification, whereby listeners are not asked to simply identify individual phonemes, but to situate those phonemic contrasts at the word-level in order to differentiate meaning.

However, having a musical background did not appear to be particularly advantageous for TM, as there was no significant difference between these participants and TNM in overall word learning attainment level. This asymmetry was further highlighted by the musical aptitude and tone identification regression results, as higher musical aptitude and tone identification scores both significantly predicted word learning success for the English listeners, but not for Thai listeners. These findings point to a differential in relevance of musicality depending on linguistic background. These results can be interpreted as consistent with previous research on non-native segmental perception, where musicianship was found to be beneficial in cases where listeners are dealing with unfamiliar and difficult L2 contrasts (Slevc and Miyake, 2006), but not in contexts where listeners are dealing with L2 phonemes that are familiar or even linguistically relevant in their L1 (Delogu et al., 2010). With respect to the present study, one possible explanation is that musical training may not bear as much influence on tone word identification for participants with an existing native tone language background. Thai listeners already have native experience with using pitch lexically, which may account for why no significant performance accuracy discrepancy between TM and TNM was found. In particular, Thai listeners are familiar with the process of mapping a change in pitch to a lexical semantic change, a mechanism established during first language acquisition for tone language speakers. It is conceivable then that musicianship does not develop or enhance this process any further. Their experience with this mapping process in their native language may have transferred to a non-native language, facilitating their ability to connect pitch to meaning in newly forming lexical representations. On the other hand, EM did not possess experience with using tone patterns to differentiate word meaning. Thus, their musical pitch expertise was drawn upon to enhance their ability to utilize linguistic pitch in a higher-level linguistic context, consistent with the notion that musical training involves “active engagement with musical sounds and the connection of ‘sound’ to ‘meaning’” (Kraus and Chandrasekaran, 2010, p. 600).

Regarding the tone identification task, where no higher level semantic processing was involved, the results of the present study confirmed our prediction that musical experience would have a significant impact, regardless of L1 background, as both EM and TM had higher accuracy rates across tests, as compared to their respective non-musician counterparts. This corroborates previous behavioral research on music and tone identification (e.g., Alexander et al., 2005; Delogu et al., 2010; Gottfried, 2007), and extends it to musicians with a tone language background. The transferability of musical pitch skills to the linguistic domain counters the notion that music and language are dissociated (e.g., Bever, 1975; Peretz and Coltheart, 2003) and suggests that more domain-general mechanisms may be employed in processing, at least at the phonemic level. Interestingly, EM also had significantly higher accuracy rates than TM. One possible explanation is that Thai listeners in general were suffering from a larger degree of interference from their L1 tone system than the English listeners. While TNM did not differ significantly from ENM, there was a slight trend for lower tone identification scores in the pre-training task for TNM relative to ENM; though, ENM’s relatively poor pitch acuity may have prevented the gap between TNM and ENM from being significantly larger. The enhanced pitch discrimination skills of TM allowed them to perform significantly better than TNM. However, interference from their L1 tone system may have inhibited their performance relative to EM, who in possessing enhanced pitch acuity and being relatively unburdened by interference from a native lexical tone system was able to perform significantly better than TM.

**C. Interaction of linguistic and musical experience**

The comparison between the performance of EM and TNM examined whether linguistic experience facilitated identification of tone words to a greater degree than musical experience or vice versa. While we predicted TNM to
outperform EM due to the linguistic nature of the task, the results revealed no significant difference in overall attainment level between these two groups. As discussed above, it is plausible that a larger degree of L1 tone interference may have inhibited the Thai listeners relative to the English listeners, as the tonal systems in the L1 and L2 differ, making it more challenging to incorporate tone information into the newly-forming lexical entries. This interference may have prevented TNM from significantly outperforming EM. However, their L1 experience with the process of utilizing pitch lexically to make pitch-to-semantic mappings may also have prevented them from performing significantly worse than EM.

More surprising is that TM did not differ significantly from ENM, with even a trend towards lower word learning proficiency relative to EM and TNM. This group was expected to outperform other groups as a result of the combination of musical and tone language experience; however, the present results indicate that the combination of these factors is much more complex than the anticipated additive effect. One speculation as to the locus of TM’s difficulty is in a conflict between what mechanisms are developed and to which dimensions of information they are attuned as a result of their musical and linguistic experience. Their musical training might tune them into attending to pitch information more actively, which might have resulted in focusing their attention on building “more elaborated percepts of the speech signal” relative to their non-musician counterparts (Besson, Chobert, and Marie, 2011, p. 3). While this was facilitative for EM, TM’s building of elaborated tone representations also required negotiating with their existing L1 tone categories. Musicianship had developed an enhanced ability to focus their attention on different acoustic dimensions; however, extracting and processing the tonal information, negotiating with the existing tone system and building pitch-to-semantic connections may have ultimately inhibited their performance relative to EM and TNM. TNM lacked the kind of enhanced pitch acuity demonstrated by TM and did not possess prior musical experience that might have focused their attention on these tone contrasts; thus, they relied more heavily on mechanisms for tone word learning developed by their linguistic experience. Rather than concentrating on actively building accurate, elaborated tone representations to incorporate into the newly forming lexical representations, TNM might have focused their attention on mapping semantic information to what tonal information they did have. Thus, it is possible that TM’s marginally lower tone word learning proficiency reflects a conflict between the type of information and strategies their musical and linguistic experience orient them towards. This could be considered in line with previous findings pointing to a dynamic relationship between aspects of listeners’ expertise, task demands and the functional significance of the sounds (Bidelman et al., 2011a).

However, both EM and TNM significantly outperformed ENM, indicating that both of these experiential factors aided tone word learning, at least at this initial stage of acquisition. These findings provide support for the idea that tone language and musical experience similarly shape the cognitive mechanisms recruited during the initial stages of tone word learning (Koelsch, Kasper, Sammler, Schulze, Gunter, and Friederici, 2004; Patel, 2003, 2008). Previous research has pointed to a convergence in music and language processing for other domains of linguistics (e.g., syntactic processing), suggesting that music and language may be dissociated at the level of representation, but overlap for cognitive processing (Patel, 2003). Neurobiological findings can also support this notion, whereby enhanced pitch-tracking accuracy in the auditory brainstem while perceiving non-native tones has been reported for both tone language listeners (Krishnan et al., 2010) and musicians (Wong et al., 2007). This indicates that the sensitivity of brainstem neurons that extract relevant pitch information is enhanced by experience-dependent subcortical mechanisms. Consistently, given that TNM and EM both outperformed ENM, these results suggest that sustained linguistic pitch experience (tone language listeners) or long-term musical pitch experience (musicians) influence the pitch processing mechanisms which aid in the acquisition of tone words for the initial learning stage.

Furthermore, tonal awareness, either inherent or trained with musical experience, has been shown to impact tone word learning success (Wong and Perrachione, 2007). The present regression results with pre-training tone identification and session 7 scores revealed that the ability to effectively identify non-native tones translated into greater word learning success for English listeners (e.g., Fig. 4). These regression results suggest that more delineated or stable non-native tone categories, as reflected by better tone identification accuracy, facilitated feature-to-word mappings for English listeners. Indeed, these findings suggest that English listeners relied on lower-level pitch abilities to help construct more stable lexical representations through the incorporation of the necessary tone information. Interestingly, the correlation failed to be significant for Thai listeners. The lack of correlation between tonal awareness with word learning may suggest that the ability to explicitly identify tone categories is not an accurate measure of tone word learning abilities for tone language speakers. Thai listeners, particularly the non-musicians, might not be as experienced with explicitly categorizing tone information, actively separating the tonal from the segmental information. Implicit tone categorization is more consistent with the nature of their experience during first language acquisition. Therefore, the tone word learning paradigm lent itself to this type of implicit internalization of tones, perhaps more in line with their experience. Indeed, studies have revealed that behavioral performance can be dissociated from neurophysiological pitch encoding. For example, English musicians and tone language listeners were found to have stronger brainstem representations of musical pitch sequences relative to English non-musicians, but this did not translate into enhanced behavioral discrimination for the tone language listeners (Bidelman et al., 2011a). The authors concluded that whether or not encoded features are utilized is dependent on a confluence of factors, including task demands and the nature of the listener’s experience. In the current study, the nature of the ‘Thai listeners’ experience might have aligned better with the demands of the tone word learning task than with the tone identification
task, such that in the former task, they were able to more effectively employ the encoded features in conjunction with their L1 mechanisms, such as pitch-to-semantic mapping processes.

D. General discussion

While previous research has primarily focused on how musical and linguistic experience influence non-native tone perception separately (e.g., Alexander et al., 2005; Francis et al., 2008; Gottfried, 2007; Wayland and Guion, 2004; Wong and Perrachione, 2007), the current results provide some insight into the relative and combined influences of both linguistic and musical factors in learning to identify non-native tone words. Furthermore, the present study investigated these factors in the context of both tone identification and tone word learning, allowing for the examination of these factors’ input at lower and higher-level contexts, extending previous literature on the relationship between phonetic and lexical recognition (Curtin et al., 1998; Weber and Cutler, 2004).

The results of the current study illustrate that the influence of linguistic and musical experience may be modulated by the level of processing (lower versus higher level contexts), and that different aspects of these factors are utilized in differing contexts and stages of learning. Indeed, the effects of one factor do not appear to be uniform throughout the different levels and stages (for example, uniformly advantageous or inhibitory). The present study found that tone language experience was not necessarily facilitative at the level of tone identification, as demonstrated by the pre-/post-training tone identification scores, consistent with previous findings (e.g., Francis et al., 2008). At this lower-level context (tone identification), L1 phonetic inventories and perceptual cue weightings appeared to play a role, manifesting in differing tonal accuracy and confusion patterns. These findings can be considered in the context of L2 speech learning models (e.g., Best and Tyler, 2007; Flege, 2007), whereby identification performance of L2 contrasts can differ depending on the relationship between the L1 and newly forming L2 categories. Indeed, group differences were observed in which tones were easier or more challenging to perceive, which may relate to differences in native tonal inventories.

Yet it was musical experience that significantly influenced overall success on tone identification, rather than prior experience with discerning linguistic pitch distinctions. This provides additional support for the notion of domain-general rather than domain-specific processing of pitch information (Zatorre and Gandour, 2008), as extensive musical pitch exposure beneficially transferred to aid linguistic tone identification (e.g., Alexander et al., 2005). These results point to more general auditory processing mechanisms, enhanced as a result of musical training, being utilized to process lower-level linguistic information.

However, there were different cognitive requirements for listeners in a higher-level linguistic context such as word learning (Strange and Shafer, 2008), which subsequently shifted what these factors contributed. Prior experience with utilizing linguistic pitch in a lexically significant manner was useful at a later stage in the learning process, allowing TNM to achieve greater word learning success than ENM. Furthermore, musical experience facilitated tone word learning for listeners without a tone language background (EM versus ENM), but no such facilitating effect was found for those with a tone language background, revealing that the combination of musicianship and tone language experience may not be additive. Musicality appeared to be a significant factor in this context only when native-language experience had not previously developed the appropriate processes involved in feature-to-word mapping. The interplay of these factors has interesting ramifications on the nature of the mechanisms involved in category identification and feature-to-word mapping. These findings suggest that linguistic and non-linguistic factors have dynamic roles in non-native perception, and that their influences may vary depending on the needs of the context. This is consistent with research suggesting that sound processing, rather than being within a modular system with domain-specific components for processing music and speech separately, occurs within an adaptive network, influenced by factors such as the input and task type, attentional considerations and the particular experience of the listener (Bidelman et al., 2011a; Marie, Kujala and Besson, 2012; Tervaniemi et al., 2009). Indeed, the present results are in line with Zatorre and Gandour (2008), who suggest that a more complete picture of tonal processing will arise from considering the influences of both general auditory processes along with those relating to linguistic knowledge. The present research provides support for an integrated model of tonal processing in a higher-linguistic context, as both language-specific experience (from L1 tonal background) and general acoustic tonal experience (from musical training) had an effect on tone word learning.

V. CONCLUDING REMARKS

The goal of the present research was to examine the relative and combined influence of linguistic and musical experience on Cantonese word learning and tone perception. The results of the current study provide evidence that musical and tone language experience both shape the cognitive mechanisms involved in pitch processing, at both lower (phonetic identification) and higher linguistic levels (word learning). However, this research also highlights the fact that the influence of and relationship between linguistic and musical experience are by no means straightforward, with issues such as linguistic context as well as other lower-level sensory and higher-level cognitive processes interacting with these factors and affecting their relative contributions during learning. Thus, further studies are needed to take into account these additional factors, such as linguistic context, intelligence, language aptitude and learning style, on tone word learning. In addition, as learning is a dynamic process, it remains for future research to investigate what role musical and linguistics factors play at different learning stages. It is conceivable that as the L2 develops, the more domain-general cognitive mechanisms utilized for the initial learning stage in this study may give way to more domain-specific processes as linguistic contexts increase in size (e.g., phrasal,
sentential, or discourse levels). Such research may further contribute to our understanding of L2 suprasegmental perception and the effect of experiential factors at different stages of learning.

ACKNOWLEDGEMENTS

We would like to thank Akkaporn Cooper, David Potter, Amy Lee, Pattama Visuttipitkul, and Tachai Luangvaranunt for their assistance in stimulus development, recruiting and data collection on this project. We also thank Ann Bradlow, Matt Goldrick, and two anonymous reviewers for their insightful comments and valuable suggestions. Special thanks to Tom Loughlin from the Department of Statistics at Simon Fraser University for assistance in statistical analysis. This study was funded in part by a research grant from the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery Grant - 312457-2006). Portions of this research were presented at the 5th International Conference on Speech Prosody, Chicago, IL, 2010, Interspeech 2010, Makuhari, Japan, 2010, and the 2nd PanAmerican/Iberian Meeting on Acoustics, Cancun, Mexico, 2010.

To clarify the distinction between “lexical tone” and “tone word,” tone words refer to lexical items minimally distinguished by phonemic tones. Tone word identification involves identifying the meaning of different lexical items. Tone identification involves the identification of phonemic tone categories.


