EFFECTS OF LINGUISTIC AND MUSICAL EXPERIENCE ON CANTONESE TONE WORD LEARNING

by

Angela Cooper
B.A., Simon Fraser University, 2007

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

In the
Department of Linguistics

SIMON FRASER UNIVERSITY

Spring 2010

APPROVAL

Name: **Angela Cooper**

Degree: **Master of Arts**

Title of Thesis: **Effects of linguistic and musical experience on Cantonese tone word learning**

Examining Committee:

Chair: **Dr. Nancy Hedberg**
Associate Professor of Linguistics,
Simon Fraser University

Dr. Yue Wang
Associate Professor of Linguistics,
Simon Fraser University,
Senior Supervisor

Dr. Murray Munro
Professor of Linguistics,
Simon Fraser University,
Supervisor

Dr. Dean Mellow
Associate Professor of Linguistics,
Simon Fraser University,
Supervisor

Dr. Valter Ciocca
Professor of Audiology and Speech Sciences,
University of British Columbia,
External Examiner

Date Defended/Approved:

ABSTRACT

Adult non-native perception is subject to influence from a variety of factors, including linguistic experience as well as other cognitive functions such as musical experience. The present research examines how these two factors influence non-native tone perception and word learning. Native Thai and English listeners, subdivided into musician and non-musician groups, engaged in a perceptual training program. They learned words distinguished by five Cantonese tones during training, also completing pre- and post-training lexical tone identification tasks. The findings suggest that musical experience or a tone language background lead to significantly better word learning proficiency over non-musically trained non-tone language listeners. Furthermore, language background appears to influence the relevance of musicality, as the combination of tone language and musical background did not provide an advantage for learners. These results point to shared processing mechanisms of music and language, both at the level of tone identification and at the word learning stage.

Keywords: lexical tone perception; word learning; Cantonese; musical experience; linguistic experience; second language acquisition; perceptual training

To my family, and to David

“Music is the universal language of mankind.”

— HENRY WADSWORTH LONGFELLOW

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1: INTRODUCTION

While the challenges that face adult second language learners have been well documented (e.g. Best & Tyler, 2007; Flege, 1995; Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 2002), there is evidence to suggest that a variety of experiential factors can influence how difficult certain non-native phonemic contrasts are to acquire. Linguistic background is considered a major factor in determining learner performance, including how the nature of the native phonetic inventory relates to the second language inventory (e.g. Best, McRoberts, & Goodell, 2001; Flege, 1995), as well as the amount of second language use (e.g. Flege, Bohn & Jang, 1997) and the age of learning (e.g. Flege, Munro & MacKay, 1995). Moreover, a variety of non-linguistic factors, including neurological maturation, learner motivation and attitude, are also purported to impact learning (e.g. Gardner, Lalonde, Moorcroft, 1985; Lenneberg, 1967). One particular extralinguistic factor, musical experience, has also been shown to contribute to second language speech learning (e.g. Slevc & Miyake, 2006), as enhanced auditory acuity developed as a product of long-term auditory training may aid listeners in discerning difficult non-native distinctions.

These factors influence not only non-native segmental but also suprasegmental perception (e.g. Alexander, Wong, & Bradlow, 2005; Delogu, Lampis, & Belardinelli, 2009). Many of the world's languages, such as Mandarin 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. This can be a challenging element to

acquire for second language learners, particularly for listeners whose native languages are non-tonal (e.g. English). Musical background has been shown to be particularly influential on suprasegmental perception, due in large part to the overlap between language and music in the relevant acoustic features, such as fundamental frequency and duration. Non-native tone studies have largely examined the influence of linguistic and musical experience separately (Alexander et al., 2005; Wayland & Guion, 2004), and there is indeed a paucity of research investigating the interaction of these factors on non-native perception. Examining the interaction of these factors may provide insight into the nature of the cognitive mechanisms employed in processing non-native contrasts and how these mechanisms are shaped by different experiences. Furthermore, the majority of research has focused on the influence of experience on the perception and learning of individual speech contrasts, 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, Goad & Pater, 1998; Wong & Perrachione, 2007).

Accordingly, the aim of this thesis was to investigate the relative and combined influences of linguistic and musical experience on non-native tone acquisition and word learning in Cantonese. In order to tease apart these influences, the present study involved participants with tone (Thai) and non-tone (English) language backgrounds further subdivided into musician and non-musician groups. To investigate whether either of these factors facilitates tone word acquisition to a greater degree than the other, a perceptual training program was employed, where listeners learned to map semantic information onto novel vocabulary words minimally distinguished by Cantonese tones. Additionally, a pre- and post-training lexical tone identification task was administered to examine

whether the ability to identify non-native tones would translate into word learning proficiency. This task also allowed us to investigate whether word learning would transfer to improved lexical tone identification accuracy. The findings of this research will provide some insight into how music and language are processed, as well as the nature of the cognitive mechanisms involved in category identification and feature-to-word mapping.

1.1 Organization of the Thesis

Chapter 2 examines previous literature on non-native perception, including a review of second language speech learning theories and empirical findings as well as a discussion of previous segmental and suprasegmental training studies. Prior research investigating the influence of linguistic and musical experience on non-native segmental and suprasegmental perception, as well as word learning, is also reviewed. This chapter concludes with an outline of the current research, research questions and hypotheses. In Chapter 3, the methodology of the experiment is outlined, with details on the participant groups, the nature of the stimuli used and procedures for the different tasks employed. Chapter 4 compares the results of the identification and word learning tasks for the different groups, and the various analyses conducted. Such analyses include percent correct tone and word identification, confusion matrices of tones, error type of tone word identification, and correlations of pre-training identification and musical aptitude scores with tone word scores. Chapter 5 discusses the implications of the current findings in the context of previous literature and theories. The first two sections of this chapter interpret the results for each task separately. The general discussion section examines how the results of these different tasks come together to shed light on the role of the main factors

in second language learning. Chapter 6 concludes the thesis with an overview of the major findings as well as future directions for this line of research.

2: LITERATURE REVIEW

Section 2.1 reviews the two main second language (L2) speech learning theories, followed by a discussion in Section 2.2 of factors that influence L2 phonetic and phonological learning, with a specific focus on the role of linguistic background and musical experience. Section 2.3 addresses studies on the training of non-native segmental and suprasegmental contrasts. Additionally, Section 2.3.3 reviews recent studies discussing the progression from non-native feature acquisition to higher-level learning, such as lexical identification. Section 2.4 introduces tonal systems and details the specific nature of the Thai, Cantonese and English systems. Finally, Section 2.5 summarizes the relevant issues addressed in the literature, and Section 2.6 outlines the central research questions and hypotheses for the present study.

2.1 L2 speech learning & theories

Infants acquiring their native language are open and receptive to a wide range of phonetic contrasts from other languages. However, after extensive native-language exposure, this perceptual receptivity declines to the point that it becomes significantly more difficult for adult learners to perceive and produce these foreign contrasts (Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 2002). Previous research seeking to elucidate this decline in perception and production can been divided into “nature” versus “nurture” perspectives (e.g. Flege, 1988; Lenneberg, 1967; Penfield, 1965). The latter perspective attributes these changes to the role of experiential factors, and the former points to biological changes during puberty, particularly neurological maturation. With

respect to the "nature" debate, studies have posited that neurological maturation results in a decline in neural plasticity, which subsequently leads to a diminished capacity for acquiring and producing new phonetic categories (Lenneberg, 1967; Penfield, 1965). The neurological changes that accompany puberty are said to delimit a critical period in speech learning, after which point attaining fluency in a new language becomes highly unlikely (e.g. Lenneberg, 1967; Patkowski, 1990).

Concerning the argument for experiential effects, several hypotheses to explain the decline in adult performance have been proposed, such as insufficient phonetic input or a lack of motivation to socio-psychological reasons for retaining a foreign accent (Flege, 1988). However, several theoretical models of L2 learning have also been posited, notably Flege's Speech Learning Model (SLM; 1995, 2007) and Best's Perceptual Assimilation Model (PAM; 1994, 1995). The former model is primarily concerned with experienced listeners (e.g. early or late bilinguals); whereas, the latter focuses on naïve listeners. The SLM posits that the mechanisms used to acquire L1 phonetic categories are retained in adulthood and are applicable in L2 learning, contra the aforementioned claims for a critical period. It also proposes that the phonetic items comprising an individual's L1 and L2 phonetic inventories exist in a "common phonological space" and can therefore be mutually influential on each other (Flege, 1995; p. 239). The SLM hypothesizes that phonetic distinctions are more likely to be detected as the perceived phonetic disparity between a new sound and its closest L1 counterpart increases. However, this ability to discern phonetic differences between L1 and L2 sounds and novel L2 categories is predicted to decrease as age of learning increases. Furthermore, the SLM posits that the development of an L2 category can be inhibited by a mechanism

termed “equivalence classification” (Flege, 1995; p. 239), whereby phonetically-similar L1 and L2 categories will assimilate and subsequently approximate each other in production (e.g. MacKay, Flege, Piske, & Schirru, 2001). Several studies have supported this model of L2 learning, in that a variety of experiential factors including amount of L2 experience (Flege, Bohn & Jang, 1997; Flege, Frieda & Nozawa, 1997; Piske, Flege, MacKay & Meador, 2001; Trofimovich & Baker, 2006) and age of learning (Flege, Munro & MacKay, 1995; Mayo & Florentine, 1997; Piske, MacKay & Flege, 2001) have a significant influence on how non-native sounds are perceived and produced. Additionally, studies have also pointed to the validity of the SLM’s claim that the degree of perceived phonetic dissimilarity can influence how successful L2 learners are in forming L2 categories (Aoyama, Flege, Guion, Akahane-Yamada & Yamada, 2004; Flege, Munro & Fox, 1994).

While the SLM deals primarily with experienced L2 learners, Best’s Perceptual Assimilation Model (1995) focuses on naïve non-native listeners. Best and Tyler (2007) propose extensions to PAM to include L2 learners; however, given that the current study focuses on naïve listeners, PAM’s original postulates and hypotheses will be discussed (Best, 1995). It posits that varying degrees of difficulty in discriminating non-native phonemes results from the perceptual relatedness of native and non-native phonetic categories. The authors argue that listeners will perceptually assimilate non-native sounds to the native phonemes that share the most articulatory similarities, and that their discrimination performance can be predicted based on how well these sounds assimilate. For instance, two non-native phonemes perceived as exemplary tokens from two distinct native phonemic categories (“Two Category assimilation”) are predicted to be easier to

discriminate than when they are perceived as acceptable exemplars from a single native phoneme category ("Single Category assimilation"). If a sound is perceived as being so phonetically disparate from any native sound, it is considered "Non-Assimilable" and perceived as a non-speech sound, with good discrimination accuracy predicted.

Best et al. (2001) provided evidence supporting this model in their investigation of American English listeners' assimilation of Zulu consonants. Three pairs of contrasts were utilized, including voiceless versus voiced lateral fricatives, voiceless aspirated versus ejective velar stops and plosive versus implosive bilabial stops. The authors anticipated that the lateral fricatives would show "Two Category" assimilation, due its use of a place of articulation foreign to English listeners, and be the easiest to discriminate. On the other hand, the bilabial stops were hypothesized to demonstrate "Single Category" assimilation, as they are both similar exemplars of the English /b/. These predictions were borne out in their data, with discrimination for the "Two Category" case being better than the "Single Category" case. Additional studies examining non-native vowel (Best, Halle, Bohn & Faber, 2003), approximant (Best & Strange, 1992), dental stop cluster (Hallé & Best, 2007) and lexical tone perception (So, 2005) have reported findings supporting the predictions made by the PAM model.

In summary, SLM posits that phonetic category representations dynamically change over the course of a person's life, modifying to become more similar or dissimilar from new L2 categories as they are introduced. PAM proposes a taxonomy of discrimination success for non-native contrasts based on the degree of their articulatory similarity to L1 phonetic contrasts. Both models highlight L1 experience as being a significant factor on the perception and production of foreign sounds.

2.2 Factors influencing adult second language phonetic/phonological learning

As illustrated in Section 2.1, it is evident that a variety of experiential factors such as age of L2 learning, amount of L2 exposure, and the acoustic/articulatory characteristics of the L1 and L2 inventories can significantly impact how efficiently and effectively non-native contrasts are acquired. This section will delve into more detail on the role of language background on non-native perception and production. In addition to linguistic experience, we will also examine how a non-linguistic factor, musical experience, accounts for variation in non-native performance.

2.2.1 Linguistic Experience

While it is widely-accepted that adult listeners have difficulty acquiring non-native contrasts, it is not the case that all contrasts are uniformly challenging for all listener groups (Flege, Bohn & Jang, 1997). As noted in Section 2.1, a learner's native phonetic systems can strongly influence the perception of novel sounds. As will be discussed in the following sections, the interaction of new phonetic structures with established ones can have facilitative or inhibitory effects on learning.

2.2.1.1 Influence on non-native segmental perception/production

The role of the native language phonological inventory on non-native perception is well documented, particularly with the classic case of English /ɹ/-/l/ perception by Japanese listeners (Best & Strange, 1992; Mochizuki, 1981). Japanese only possesses two contrastive approximants, as compared to the four-way contrast in English. This discrepancy leads to the mapping of /ɹ/ and /l/ onto the Japanese alveolar tap /ɾ/ or the labio-velar approximant /w/, causing considerable challenges for perception and

production. Interestingly, non-native perceptual difficulties can still arise even if the L1 phonological inventory does possess a phonemic /ɹ/. Hallé, Best, and Levitt (1999) demonstrated that despite the existence of an /ɹ/ category in French, the articulatory-phonetic characteristics of the French /ɹ/ are distinct enough from the English /ɹ/ to cause French listeners perceptual challenges. The authors suggest that rather than relying solely on abstract phonological categories, the phonetic distinctions between the two categories can account for these perceptual difficulties. Additionally, Weinberger (1990) found an effect of linguistic experience for Japanese and Russian learners' productions of English /ɵ/, who consistently mispronounced the segment as /s/ and /t/ respectively.

Similar findings have been demonstrated for the perception of non-native vowels. Rochet (1995) suggested that L1 phonology plays a role in perception, in that the French /y/ was perceived differently by native Portuguese and English listeners (/i/ and /u/, correspondingly). In addition, Ingram & Park (1997) found an L1 transfer effect on Korean perception of the English /e-æ/ contrast, reporting a differential in perception between older and younger Korean listeners, resulting from a phonemic merger in progress in Korean. As a whole, these studies illustrate how native language-specific consonant and vowel inventories can influence how L2 contrasts are produced and perceived.

2.2.1.2 Influence on non-native suprasegmental perception

Studies have also investigated the effect of listeners' native phonetic systems on their perception of non-native suprasegmental contrasts. In a study examining the influence of linguistic experience on lexical tone perception, Gandour (1983) reported

that perceptual discrepancies between language groups could be attributed to the relative weighting of two perceptual dimensions: F0 height and direction of change. Five language groups participated in the study, including Taiwanese, Mandarin, Cantonese, Thai and English listeners. Participants completed a paired similarity rating task on 19 lexical tones (14 contour, 5 level) compiled from the four participating tone languages. Results indicated that non-tone language listeners (English) attached less importance to the 'direction of change' dimension and gave greater weight to the 'height' dimension than the tone language groups. Of the tone language groups, Cantonese listeners found the 'height' dimension to be more critical than 'direction', which was attributed to the directional properties of native Cantonese lexical tones. Conversely, 'direction of change' appeared to be the most perceptually salient dimension for Thai listeners. The author concluded that cross-linguistic differences in speech production and perception are highly influenced by extensive exposure within a given linguistic environment.

Schwanhäuβer (2008), examining the categorical perception of synthetic rising and falling tone continua, reported that listeners from different tone language backgrounds employed different perceptual strategies when identifying the tone contours, in that the perceptual anchor was located mid-continuum for the Mandarin and Vietnamese listeners and around the flat stimulus pairs for the Thai listeners. Schwanhäuβer posited that the nature of listeners' tonal inventories could have a profound influence on categorical tone perception. Similarly, in a study examining native English listeners' perception of Mandarin lexical tones, Broselow, Hurtig and Ringen (1987) suggested that the English intonational system had a direct influence on how listeners were perceiving particular tones, namely the falling tone (tone 4). The authors

argued that English declarative intonation and the Mandarin falling tone share acoustic similarities, thereby facilitating the identification of this tone in isolation as well as when it occurred in final position of a string of syllables. In the latter case, it is suggested that English listeners are more accustomed to hearing standard declarative intonation, entailing a decline in pitch at the ends of sentences. Thus, the authors suggested that a falling Mandarin tone at the end of a series of words was easier to identify for the English listeners because their native intonational pattern served as a primer, facilitating the assimilation of a sentence-final falling tone with the sentence-final rise-fall in English intonation.

Research has also suggested that it may not be just the mapping of non-native tones onto native tone categories that can influence acquisition but also just having experience with being attuned to F0 changes and distinctions in one's native language (Y. Lee, Vakoch, & Wurm, 1996; Wayland & Guion, 2004). Lee et al. (1996) tested native Mandarin, Cantonese and English listeners with a same-different discrimination task for pairs of Cantonese and Mandarin tones. Their results showed that Cantonese listeners had higher discrimination accuracy than the English group on Mandarin tones, and that Mandarin listeners performed with higher accuracy than the English listeners on Cantonese tones for certain conditions. The authors suggested that tone language listeners acquire broad tone discrimination skills through exposure to and familiarity with native tones. These findings were supported by Wayland and Guion (2004), who employed a 5-day training program with native Thai, English and Chinese participants learning to distinguish Thai mid and low tones. A categorical-oddity task, where listeners indicated which of three syllables had a different tone than the other two, was utilized. The authors

found a significant improvement in discrimination accuracy after training for the Chinese but not the English group. They also reported that the Chinese group significantly outperformed the English group, before and after training, at discriminating the tones and locating the position of the oddball tone. They suggested that native language experience with detecting word-level F0 changes could have positively transferred to aid non-native tone discrimination. Moreover, there is also neurological evidence to support this behavioural difference between tone and non-tone language listeners, whereby tone language listeners were reported to have developed more sensitive brainstem mechanisms for pitch variations (revealed by pitch-tracking accuracy and pitch strength patterns) as compared to non-tone language listeners (Krishnan, Gandour, & Bidelman, 2010; Krishnan, Swaminathan, & Gandour, 2008).

In contrast, other studies have suggested that native tone language experience does not necessarily predict greater success over those with non-tone language backgrounds in perceiving non-native tones (Francis, Ciocca, Ma, & Fenn, 2008; So, 2006a; Wang, 2006). Instead, they posit that the nature of the native tonal inventories and how they interact with incoming non-native tones can better explain differences in cross-linguistic tone perception. Francis et al. (2008) tested native Mandarin, Cantonese and English listeners on Cantonese tone perception, utilizing a pre-/post-test design with perceptual training. Participants identified the tones of words in carrier sentences and performed a difference rating task. Unlike other studies such as Wayland & 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 were evident in which tones were more challenging. Consistent with the notion that native category representations can have a significant influence on non-native perception, Mandarin listeners' performance was best on the three Cantonese tones that have similar counterparts in Mandarin.

Research has also proposed that the relative significance (functional load) of a particular suprasegmental contrast in one's native language can influence non-native perception. McAllister, Flege and Piske (2002) tested Estonian, English and Spanish listeners on their perception and production of Swedish vowel quantity, finding that performance accuracy decreased as a function of the role of temporal distinctions in the native language. Estonian listeners, whose native language employs contrastive quantity distinctions, were the most successful, followed by the English and Spanish groups. The authors posit that the differential in performance between the English and Spanish listeners could be attributed to the fact that English listeners are more sensitive to temporal cues than the Spanish, as some English listeners are capable of identifying vowels based on length. Their results support the notion that the relative importance of a feature in one's native language can affect how easily it is perceived and produced in an L2.

Neurological studies have also reported language-dependent variation in non-native lexical tone processing, particularly with respect to the degree of linguistic significance or function (Gandour et al., 2000; Gandour, Wong, Lowe, & Dzemidzic, 2002; Wang, Behne, Jongman, & Sereno, 2004). Linguistic information is primarily processed in the left hemisphere (e.g. Studdert-Kennedy & Shankweiler, 1970), including linguistically-significant tonal contrasts (Van Lancker & Fromkin, 1973; Wang,

Jongman, & Sereno, 2001). However for non-native listeners, Gandour et al. (2000, 2002) noted that even if pitch is linguistically relevant in their native language, participants did not show left hemisphere dominance when perceiving non-native tones. Similarly, Wang et al. (2004) attributed cross-linguistic differences in hemispheric lateralization of Mandarin tones to linguistic relevance. Native Mandarin and early English-Mandarin bilinguals demonstrated left hemisphere superiority, as these tones were linguistically meaningful and were thus processed similarly to other linguistic information in the left hemisphere. On the other hand, English and Norwegian (whose language makes use of phonemic pitch distinctions) listeners showed bilateral processing, as these tones did not provide any linguistic information for these listeners. These studies point to a dynamic range of lexical tone processing patterns that are dependent on linguistic experience.

Taken together, previous research has highlighted the profound impact native language experience has on non-native segmental and suprasegmental perception and production. Research has indicated that the L1 phonological inventory can influence how L2 segments are perceived and produced (e.g. Best & Strange, 1992; Hallé et. al, 1999; Rochet, 1995), as the ease with which L2 phonemic categories are formed can be affected by how perceptually similar or dissimilar they are from L1 categories. Linguistic experience has also been shown to play a role in the perception of suprasegmentals. Gandour (1983) proposed that cross-linguistic differences result from experience-dependent weightings of certain perceptual dimensions, such as F0 height and direction. Studies have posited that tone language experience develops better tone discrimination skills that can transfer to new tone languages (Lee et al., 1996; Wayland & Guion, 2004).

Others have also suggested that it is the specific nature of listeners' native tonal inventories and how L1 tone categories map onto non-natives tones that ultimately influences non-native perception (Francis et al., 2008), and that a native tone background can even hinder perception in some cases (So, 2006; Wang, 2006).

2.2.2 Musical experience

Numerous studies have pointed to a link between language and music (e.g. Besson & Schön, 2001; Patel, 2008). On a broad level, Besson and Schön (2001) pointed to the similarities in how both music and language are structured. Music conforms to two main structural criteria: rhythmic and temporal ratios that define a given piece of music by the "formalized segmentation of time" (2001, p. 235), as well as the organization of discrete successive pitch levels (scales). Correspondingly, language can be framed in comparable terms. It is similarly comprised of rhythmic and temporal patterns (e.g. stress versus syllable timing). Furthermore, musical notes can be considered analogous to phonemes, as both require the segmentation of an auditory stream into discrete units. Both language and music qualify as rule-based systems where fundamental units (e.g. notes or phonemes) are arranged into higher-level hierarchical structures.

Thus, it is not surprising then that studies have suggested language and music may have shared cognitive faculties, and that crucial language areas in the brain are recruited during the processing of music (Besson & Schön, 2001; Levitin & Menon, 2003). This may account for why verbal memory was also found to be better in musicians than non-musicians (Chan, Ho, & Cheung, 1998; Brandler & Rammsayer, 2003). This close connection between language and music has led to a growing body of research addressing the effect of musical training on non-native perception and second language learning (e.g.

Milovanov, Pietilä, Tervaniemi, & Esquef, 2010; Milovanov, Tervaniemi, & Gustafsson, 2004; Slevc and Miyake, 2006; Wong & Perrachione, 2007).

2.2.2.1 Influence on non-native segmental perception/production

Slevc & Miyake (2006) rigorously examined the relationship between musical ability and multiple facets of second language learning, including perceptive and productive phonology, syntax and lexical knowledge. Native Japanese speakers learning English were tested on their ability to discern pronunciation errors, their level of accentedness in producing English words and passages, their ability to detect syntactically well-formed sentences, as well as several other linguistic and musical tasks. Their level of musical proficiency was determined by Wing's *Measures of Musical Talent,* comprised of chord analysis, pitch change and tonal memory sections. The researchers found a significant correlation between musical ability and proficiency in L2 phonology (both receptive and productive). The authors concluded that these results demonstrate the positive influence of musical ability on L2 phonological proficiency. Similar findings have also been reported for native Finnish adults and teenagers learning English (Milovanov et al., 2004, 2010). A negative correlation between musical aptitude scores and pronunciation errors was found for Finnish teenagers and adults, whereby those who scored higher on musical aptitude tests made fewer errors in pronunciation. Finnish teenagers with higher musical aptitude scores also had fewer errors discriminating between non-native phonemes (e.g. /ʃɪp/ versus /ʃip/). These studies support the notion that musical proficiency and L2 learning are interconnected.

2.2.2.2 Influence on non-native suprasegmental perception/production

Moreover, the relationship between musical experience and proficiency with second-language suprasegmentals, particularly lexical tone, has been extensively documented (Alexander, P. C. Wong, & Bradlow, 2005; Delogu, Lampis, & Belardinelli, 2006, 2009; Gottfried, 2007; C. Lee & Hung, 2008; Schwanhäuβer, 2008). Given that both music and tone languages employ significant pitch modulations, a logical prediction would be that proficiency in one would precipitate proficiency in the other.

Alexander et al. (2005) tested native English musicians and non-musicians on their perception of lexical tones on monosyllabic Mandarin words. Participants were asked to identify the tones in a two-alternative forced choice identification task and to discriminate between pairs of tones (same or different). A significant difference in accuracy was found between the two groups, as the musicians performed with greater accuracy and quicker reaction times than the non-musicians on both tasks. From these findings, the researchers posited that musical pitch processing aptitude may transfer to the processing of linguistic pitch. Additionally, Gottfried (2007), testing native English music and non-music majors, also found an advantage of musicianship for identification and discrimination of Mandarin tones, as music majors performed with higher accuracy than the non-music majors. A native Mandarin speaker also rated their tone productions as being significantly better than the non-music majors' productions. These results highlight how musical experience can facilitate perception and production of non-native lexical tones.

Delogu et al. (2006, 2009) extended these findings to native Italian speakers. Participants completed Wing's *Standardized Tests of Musical Intelligence* and were

classified into three groups based on their scores (High, Medium or Low). Their tonal task consisted of listening to pairs of Mandarin syllable lists and indicating whether the lists were identical, or if there were tonal or phonological differences. Results showed that listeners categorized as High had significantly higher accuracy at identifying tonal variations than the Medium and Low groups; however, no group difference was found for phonological variation detection. They propose a positive music-to-language transfer effect, whereby musical proficiency selectively enhances linguistic pitch perception.

Furthermore, there is a growing body of evidence pointing to the neurological influence of long-term exposure to musical pitch and its effect on linguistic pitch processing. Wong, Skoe, Russo, Dees and Kraus (2007) reported 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 usage may alter fundamental sensory circuitry. Correspondingly, Chandrasekaran, Krishnan, and Gandour (2009) found that Mandarin tone homologues presented in a non-speech context produced greater mismatch-negativity (MMN) responses in English musicians and native Mandarin listeners as compared to English non-musicians, suggesting that pitch experience is domain-general and can therefore facilitate both linguistic and non-linguistic pitch processing.

Finally, Wong and Perrachione (2007) also reported that musical experience was a major factor in achieving tone word learning success, with seven out of the nine successful learners in the training program being amateur musicians. Similar to Alexander et al. (2005) and Gottfried (2007), the researchers found that musically-trained

participants performed significantly more accurately on the pitch pattern identification task, which was shown to be a predictor of overall word learning success.

It should be noted that the definition of musicality is by no means consistent across studies, and the relationship between musical experience and musical aptitude and its role in language perception is one that still warrants further investigation. Schellenberg (2003) suggested that individuals who have engaged in music lessons over an extended period of time will have spent hours devoted to such tasks as ear training, sight reading and timing practice, which could positively transfer to non-musical domains such as language. On the other hand, researchers have also proposed that it is actually inherent auditory processing capabilities (musical aptitude) and not necessarily musical training as such that directly contributes to enhanced second language perceptual performance (Gottfried, 2007; Schwanhäuβer, 2008). Although, it is also possible that those with higher levels of musical aptitude are more likely to continue with musical training because they may be more successful at it.

In sum, there is robust evidence to indicate that musical experience enhances auditory acuity, which can aid in the perception of difficult non-native segmental and particularly suprasegmental contrasts, such as lexical tone.

2.3 L2 segmental and suprasegmental training

As discussed in Section 2.1, over the course of native language development, the perceptual sensitivities of a given speaker become attuned to the critical acoustic characteristics of their L1, which may later cause “perceptual interference” when attempting to tune into the important cues of a foreign language (Iverson, Kuhl, Akahane-

Yamada, Diesch, Tohkura, Kettermann & Siebert, 2003, p. B55). As a result, adult learners encounter a myriad of challenges, from not only needing to acquire unfamiliar sound contrasts but also needing to map familiar meanings onto new forms. It was posited that adults had little chance of overcoming these challenges to acquire difficult non-native contrasts successfully because they had passed the critical period of speech learning (e.g. Lenneberg, 1967; Scovel, 1988). However, as this section will discuss, a multitude of studies have since demonstrated that adults' perceptual systems retain a certain level of plasticity.

2.3.1 Learning non-native segmental contrasts

Despite the steep learning curve facing most non-native listeners, previous studies have demonstrated that adults are capable of improving their ability to distinguish non-native segments, even after only a brief period of laboratory training. Logan, Lively and Pisoni (1991) trained and tested native Japanese listeners on distinguishing the notoriously challenging English /ɹ/-/l/ contrast. The source of this difficulty for learners likely arises from the nature of the phonemic inventory in Japanese, which contains two contrastive approximants (/j/, /w/); whereas, English possesses four approximants (/ɹ/, /l/, /w/, /j/). However, after three weeks of training using high-variability stimuli, which included the use of multiple talkers and a variety of phonetic contexts, Japanese participants significantly improved in their ability to identify English /ɹ/ and /l/. Subsequent studies have reported similar findings for this particular contrast (e.g. Bradlow, Pisoni, Yamada, & Tohkura, 1997; Hardison, 2005), while also demonstrating that listeners were capable of generalizing their training to a novel talker and novel words (Yamada, 1993), and were able to retain their training improvements 6 months later

(Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999).

In addition to the numerous studies addressing the English /ɹ/-/l/ contrast, additional studies have been conducted investigating the acquisition of other non-native segmental contrasts and have shown similar learning results. Jamieson and Morosan (1986) reported that even after only 90 minutes of training, there was a substantive improvement for Canadian French listeners discriminating and identifying English /θ/ and /ð/. Pruitt (1995) noted that both Japanese and English listeners find the Hindi dental and retroflex consonants challenging to distinguish. However, after 12 training sessions, listeners from both groups showed substantial improvement from pre-test to post-test, while also generalizing to new speakers and contexts. Evidence of the effectiveness of laboratory training has also been found for Voice Onset Time (McClaskey, Pisoni, & Carrell, 1983; Pisoni, Aslin, Perey, & Hennessy, 1982) and Korean voicing contrasts (Francis & Nusbaum, 2002). Taken together, these studies demonstrate that adult listeners' perceptual systems retain plasticity, as they are capable of re-tuning to discern relevant acoustic dimensions of non-native contrasts after focused laboratory training.

2.3.2 Learning non-native suprasegmental contrasts

Wang, Spence, Jongman and Sereno (1999) demonstrated that the effectiveness of laboratory training on improving perception of non-native segmental distinctions could also be applicable in the suprasegmental domain. By utilizing high-variability procedures similar to those developed by Logan et al. (1991), the experimenters trained American English listeners to identify four Mandarin lexical tones over a period of two weeks.

Training entailed a two-alternative forced choice identification task with feedback using stimuli produced by four different talkers. Pre- and post-tests were administered to determine the effect of training on identification accuracy, as well as two generalization tasks and a long-term retention test, 6 months after training. Results illustrated that an eight-session laboratory training program was able to significantly improve non-native listeners' Mandarin tone perception from pre-test to post-test. Additionally, participants were able to generalize their training to new talkers and words, as well as demonstrate successful retention 6 months later. This perceptual training was also shown to transfer to improvements in lexical tone production (Wang, Jongman, & Sereno, 2003).

Similar findings have been reported for non-native suprasegmental training in subsequent studies on lexical tone (Francis et al., 2008; So, 2006a; Wayland & Li, 2008) and vowel length (Hirata, 2004; Hirata, Whitehurst, & Cullings, 2007). Wayland and Li (2008) reported that even a 2-day training program had an impact on native Mandarin and English listeners identifying and discriminating Thai mid and low lexical tones. Francis et al. (2008) provided additional evidence for the effect of training of a different target tone language, in that native Mandarin and English listeners similarly improved their ability to identify Cantonese lexical tones in sentence contexts after 10 days of training. Hirata (2004) found that laboratory training was also effective for learning Japanese vowel length contrasts, producing significant increases in identification accuracy from pre- to post-test for trained native English listeners in both word and sentence contexts.

Thus, there appears to be substantial evidence of the plasticity of adult perceptual systems, in that even relatively short-term laboratory training can improve the discrimination and identification accuracy of challenging non-native segmental and

suprasegmental contrasts. Additionally, researchers have reported that training can extend beyond the individual trained items, with trainees generalizing to novel talkers and words. Studies have also indicated that the perceptual gains on certain contrasts can be retained months after training has concluded.

2.3.3 From features to word learning

The majority of training studies have sought to improve listeners' perception of phonetic distinctions; however, there has been little research on whether the ability to hear non-native acoustic distinctions can be applied to broader linguistic contexts such as word learning. Two studies have examined how listeners used specific non-native contrasts in lexical identification tasks (Curtin, Goad, & Pater, 1998; Wong & Perrachione, 2007), reporting that participants were capable of using specific featural distinctions, the Thai three-way aspiration and voicing contrast and the Mandarin three-way lexical tone contrast respectively, to distinguish vocabulary items. As the design of the present study followed the procedures developed by the aforementioned research, these two papers are reviewed in detail below.

Curtin et al. (1998) investigated native French and English listeners' perception of the Thai three-way aspiration and voicing contrast (i.e. voiceless aspirated, voiceless unaspirated and voiced unaspirated) and their ability to use these contrasts to identify different lexical items. It should be noted that English possesses these distinctions allophonically, but only phonemically contrasts voiced and voiceless sounds. French only possesses a voicing distinction and does not utilize aspiration phonemically or allophonically. Six minimal triads were produced by four Thai speakers, and each of the 18 words were paired with a picture representation of their meaning. An individual

training session presented learners with six sets of three segmentally-distinct but semantically-related words. Each word was presented aurally, and its associated picture was shown on the computer screen. Quizzes with feedback were provided after completing each set of words; however, they were never tested directly on minimal pairs during training, in order to observe the acquisition of these contrasts devoid of any explicit instruction. By the end of each training session, the full set of 18 words was tested. The researchers reported that training was effective at enabling both groups to distinguish lexical items by the end of the program, with identification accuracy above 98% on day 3 of training.

Two days of training (days 1 and 3) and two days of testing (days 2 and 4) were conducted, as well as an additional testing day one week later to determine long-term retention. Two tasks were administered on each testing day: a 'Minimal Pair' task and an ABX task. For the 'Minimal Pair' task, participants were tested on the words and their associated meanings learned during training. They were shown sets of three pictures, including two pictures of a minimal pair and one foil picture of a phonetically-distinct word (e.g. [don] 'boat', [ton] 'tree' and [bàk] 'pineapple'). Minimal pairs were constructed such that participants heard all possible combinations of aspiration/voicing type. Upon hearing an aurally-presented stimulus, they would indicate which picture corresponded to the meaning of that word from the set of three pictures. The ABX task consisted of listening to a series of three words: A and B words were a minimal pair, while the X word matched either A or B. Listeners selected which word (A or B) matched the X word.

For the ‘Minimal Pair’ task, listeners did not significantly improve their accuracy from day 2 to 4. They did find a discrepancy in the order of acquisition, in that both groups appeared to lexically represent the voicing contrast faster than the aspiration contrast. The ABX task, which primarily utilized listeners’ phonemic judgments, also did not find an effect of training; however, results differed based on language background. English listeners, whose native inventory contains aspiration as an allophonic distinction, performed better than the French listeners at discriminating the aspiration contrast. The authors concluded that listeners initially form lexical representations utilizing L2 features contrastive in their L1 (in this case, voicing), before lexicalizing features that are not present or phonemically relevant in their L1 phonology.

Following Curtin et al. (1998), Wong and Perrachione (2007) examined whether native English listeners are capable of utilizing non-native suprasegmental contrasts (namely lexical tones) to distinguish word meanings. Participants learned English pseudowords that were superimposed with three Mandarin lexical tones (high-level, rising, falling). A native English speaker produced six different syllables with a high pitch, which were re-synthesized in order to create the three different pitch patterns, for a total of 18 target words. A similar training design to Curtin et al. (1998) was employed, where participants learned to associate a picture with each target word. At the end of each training session, they were tested on all 18 words, and their scores indicated whether or not they had reached criterion. The researchers administered a performance-based training program, such that if participants achieved at least 95% accuracy for two consecutive sessions or did not improve by 5% for four consecutive sessions, training was terminated.

Before the start of training, participants also completed a pitch pattern identification task to gauge their pitch perception ability. Native Mandarin speakers produced five vowels with a level tone, which were re-synthesized to create the three lexical tones (similar to the training stimuli). The participants listened to each syllable and indicated the directionality of the pitch pattern by choosing from two visually-presented arrows denoting level, rising or falling patterns of movement.

Results for the training program showed that listeners were capable of learning to use non-native pitch patterns to differentiate lexical items. When attainment level was starting to be reached, participants had improved an average of 51% from their first session. Over half the participants were deemed "successful learners", having achieved 95% for two consecutive sessions. The researchers noted that the ability to perceive and identify these tones significantly contributed to success at learning the lexical items, with scores on the pitch pattern identification task being a significant predictor of attainment level. They suggested that having a pre-existing auditory aptitude can aid lexical learning.

Thus, these two studies extend the findings discussed in Section 2.3.1 and 2.3.2 by demonstrating that adult listeners are not only capable of learning to distinguish difficult non-native segmental and suprasegmental contrasts, but that they are also able to use them to identify lexical items.

2.4 Tonal systems

2.4.1 Overview

The majority of the world's languages imbue tonal variations, that is, the manipulation of fundamental frequency (F0), with meaning, such as information structure

or lexical content. The domain of tonal realization differs cross-linguistically, ranging from a single syllable to strings of sentences (Burnham & Mattock, 2007). The most lexically relevant linguistic pitch unit is lexical tone, whose domain of realization is generally the syllable. Lexical tones are phonemically contrastive, in that they are used to differentiate word meaning. Stress, which utilizes pitch amongst other acoustic dimensions such as vowel quality, duration and intensity (Fry, 1955; Lieberman, 1960), is employed by languages as a way of marking the relative prominence of syllables (Pierrehumbert & Hirschberg, 1990). Finally, intonational contours are considered tonal melodies realized at the sentential level.

For the purposes of this thesis, languages can be broadly categorized into two types based on how they utilize pitch: lexical vs. post-lexical (Yip, 2002). Nearly 70% of the world's languages are considered tone languages, which include Mandarin, Cantonese and Thai (Yip, 2002). Fundamental frequency is the primary acoustic cue associated with tone; although, most languages make use of other secondary features such as intensity, duration and voice quality (Nguyên & Edmondson, 1997; Pham, 2003; Yip, 2002). On the other hand, post-lexical tone languages such as English generally use pitch information over a broader temporal domain to convey grammatical information (e.g. declarative, interrogative), emotion, speaker and propositional attitudes as well as emphasis (Ladd, 1997; Pierrehumbert & Hirschberg, 1990). Additionally, English also employs syllable-level pitch distinctions (stress) to signal prominence and some grammatical contrasts. It should be noted that membership within one category does not preclude membership in the other, as lexical tone languages also make use of stress and intonational contours.

As discussed in Section 2.2.1.2, the nature of L1 and L2 tonal systems can have a significant influence on perception. Thus, in the following sections, the tonal systems of Cantonese, Thai and English will be described in greater detail, as these languages are directly relevant to the research questions of the present study.

2.4.2 Cantonese

The tonal inventory of Hong Kong Cantonese is comprised of six contrastive lexical tones, whose pitch contours are depicted in Figure 1. These tones include High-Level (55), High-Rising (25), Mid-Level (33), Low-Falling (21), Low-Rising (23) and Low-Level (22) (Bauer & Benedict, 1997). The numbers in parentheses refer to Chao's (1930) tone letter notation representing the overall contours of the tones based on a five-point scale, with 5 denoting the highest and 1 the lowest pitch point. Table 1 provides an example of a set of tones on one Cantonese syllable, with the English translation below each.

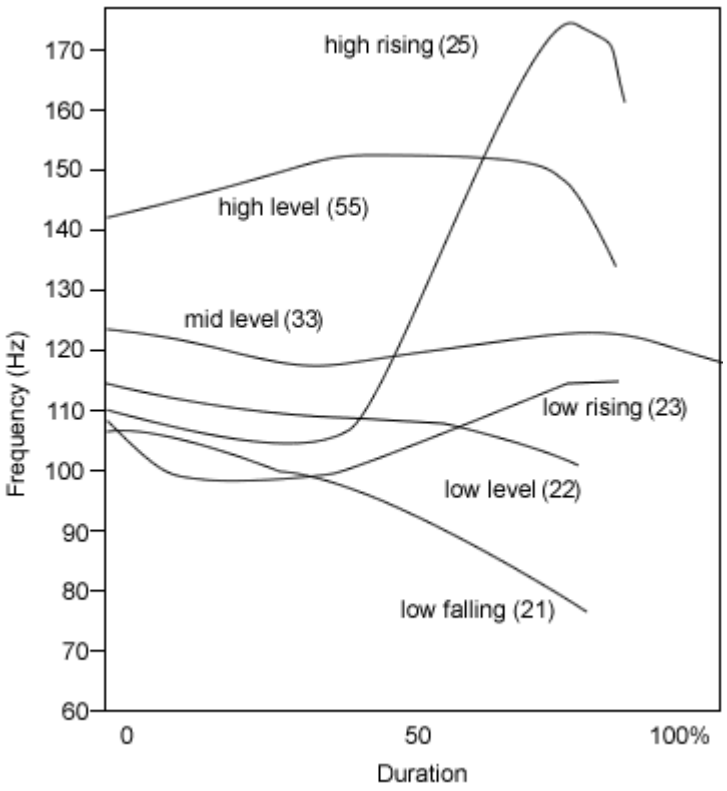


Figure 1 Cantonese tone inventory (adapted from Francis et al. (2008)). The numbers in parentheses follow Chao (1930), referring to overall tone shape derived from a 5-point scale (5=high, 1=low).

Table 1 Six-way tonal contrast in Cantonese, illustrated on syllable /ji/. The numbers in parentheses follow Chao (1930), referring to overall tone shape derived from a 5-point scale (5=high, 1=low).

High level (55)	**High rising (25)**	**Mid level (33)**	**Low falling (21)**	**Low rising (23)**	**Low level (22)**
ji clothes	*ji* chair	*ji* idea	*ji* suspicious	*ji* ear	*ji* two

Traditionally, the definition of the Cantonese tonal system included seven phonemic tones, with an additional High-Falling tone. However, most Hong Kong speakers are reported to have lost the distinction between High-Falling and High-Level and have collapsed the two categories into one, utilizing only the High-Level tone (Bauer & Benedict, 1997). The six tonal contours only occur on unchecked syllables, which include

syllables with nuclear vowels or those ending in nasals or semi-vowels. For closed syllables ending in the voiceless stops /p t k/, only three tones can occur: High, Mid and Low-Level (Kao, 1971).

While voice quality and durational differences between certain tones have been reported, relative pitch appears to be the primary perceptual cue to distinguish Cantonese tones (Vance, 1976). Additionally, pitch height may be a more critical dimension than directionality for differentiating these tones for native listeners (Vance, 1977).

2.4.3 Thai

There are five lexical tones in Standard (Bangkok) Thai, depicted in Figure 2, which include three static and two contour tones. These have been traditionally described as High, Mid, Low, Rising and Falling (Abramson, 1962). Table 2 provides a Thai syllable minimally distinguished by these tones with their English translations. While the High and Low tones are typically categorized as being static, Abramson (1979) noted that the acoustic manifestation of these tones, particularly in running speech, entails some movement, high rising and low falling respectively. Similar to Cantonese, maximal tonal differentiation is only achieved on unchecked syllables containing long vowels, dipthongs or ending in nasal consonants (Abramson, 1976). Although concurrent features such as overall amplitude and glottal tension are present, pitch is considered the primary perceptual cue for distinguishing the Thai tones, with rapid F0 movement required for accurate native perception of the contour tones (Abramson, 1975, 1978)

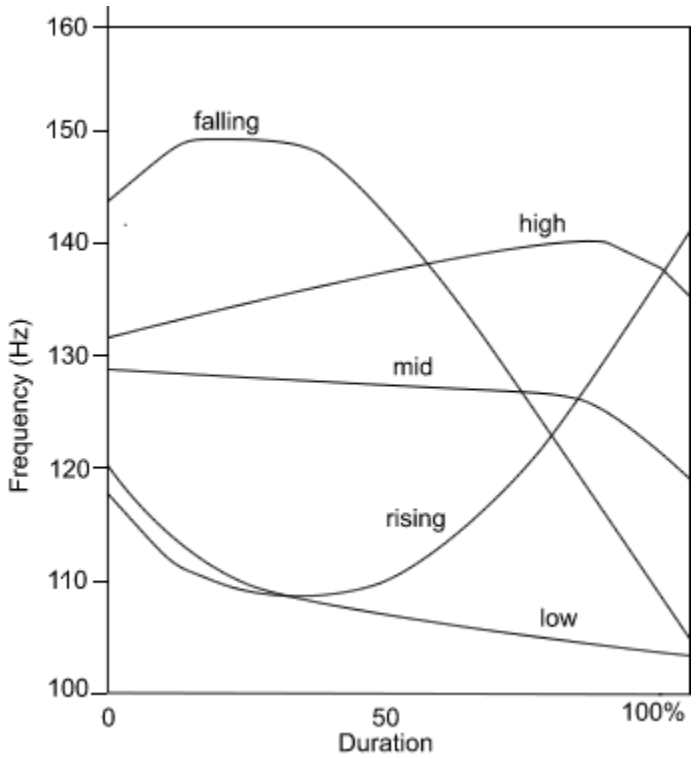


Figure 2 Thai tone inventory (adapted from Gandour, 1983, p. 152)

Table 2 Five-way tonal contrast in Thai

High	**Mid**	**Low**	**Rising**	**Falling**
kaj to scoop out	*kaj* dried sweat	*kaj* egg	*kaj* to unlock	*kaj* fever

2.4.4 English

Delineating the English intonation system is somewhat more complicated than the tonal inventories of Cantonese and Thai, as there is much debate in the literature on the precise nature of the English prosodic inventory. As discussed previously, English does not use pitch to distinguish lexical items but rather to convey different stress patterns and intonational meanings (Beckman & Pierrehumbert, 1986; Ladd, 1978; Pierrehumbert & Hirschberg, 1990). Ladd (1978) suggested that there exists an intonational lexicon for English comprised of tonal melodies individually associated with pragmatic meanings.

The exact number of these distinctive tonal melodies is the subject of some debate. Beckman and Pierrehumbert (1986) posited an intonational system composed of low (L) and high (H) tone sequences: six pitch accents (two simple tones and four complex tones), two phrasal accents and two boundary tones, the combination of which produces over twenty different intonational contours. Whether or not each individual contour possesses a unique, identifiable meaning has yet to be empirically verified. There is evidence to suggest that English listeners are capable of distinguishing several primary intonation patterns, including the rising (HL) and falling contours (LH), and that these contours are quite distinct in listeners' perceptual spaces (Grabe, Rosner, García-Albea, & Zhou, 2003). However, experimental data on the mapping of German intonation patterns to two distinct pragmatic categories relating to thematic contrast suggested that listeners are not able to reliably associate these intonation patterns with the appropriate pragmatic contexts (Braun, 2006). Thus, it is evident that the categorical nature of intonation patterns is not as clear-cut as with lexical tone contrasts.

In English, there also exists word and sentential stress, existing independent of the intonational melody, whose acoustic correlates include pitch as well as other dimensions such as vowel quality and duration (Pierrehumbert & Hirschberg, 1990). Stressed syllables are generally considered possible location sites for the intonational pitch accents that comprise intonational contours. A stressed syllable will typically have a higher F0, longer duration, greater intensity and a more fully articulated vowel than its unstressed counterpart (Fry, 1955). Stress, unlike lexical tone, is not paradigmatic, but rather forms syntagmatic relations (Gussenhoven, 2004; Szwedek, 1986). Phrasal stress patterns vary depending on information structure, as the locus of nuclear (main) stress is contingent on

which word needs to be emphasized (e.g. “John never liked the OPERA” versus “JOHN never liked the opera”). Stress is perhaps closer to lexical tone than intonation, in that they both share similar domains of realization (i.e. syllable). However, their respective linguistic functions are quite different. As discussed earlier, tone is utilized to make lexical distinctions, and thus carries a very high functional load (Surendran & Levow, 2004); whereas, stress carries a lower functional load, in that the patterns of prominence are used to convey largely pragmatic information (e.g. signaling “new” information in the discourse) and grammatical contrasts, though only for a restricted pairs of words (Cutler, 1986).

In sum, the three prosodic systems involved in the present research are Cantonese, Thai and English. Cantonese and Thai are both considered tone languages that employ phonemically contrastive pitch distinctions. Cantonese possesses six distinct lexical tones (3 level, 3 contour), and Thai has five tones (3 level, 2 contour). English, on the other hand, has been termed a post-lexical tone language (Yip, 2002), in that it uses pitch to primarily signal discourse information rather than to distinguish lexical items.

2.5 Summary

Adults face many more difficulties than infants and children when learning a second language (Werker & Tees, 2002). However, adult learners possess a breadth of experience that can potentially influence the ease with which they perceive and produce foreign contrasts. Studies have demonstrated that performance variability can be accounted for by several factors, including linguistic and musical experience (e.g. Flege, 2007; Francis et al., 2008; Wayland & Guion, 2004). Linguistic experience was found to influence the degree of perceptual difficulty of non-native contrasts based on their

acoustic and articulatory relationship to native language phonetic and phonological categories (e.g. Best et al., 2001). Several studies have suggested that having a tone language background could be advantageous when learning foreign tonal contrasts (Lee et al., 1996; Wayland & Guion, 2004). Although, other studies have posited that the mere presence of lexical tones in one's native language inventory does not necessarily entail greater proficiency in non-native tone acquisition, and that the influence of language background arises from how non-native tones map onto native tone representations (Francis et al., 2008; So, 2006a; Wang, 2006). It has also been demonstrated that musical experience can facilitate second language acquisition, with musically-trained participants consistently performing with higher overall accuracy and speed of acquisition of certain non-native segmental and particularly suprasegmental contrasts (Alexander et al., 2005; Delogu et al., 2009; Gottfried, 2007). Despite the challenges facing adult learners, and claims of neurological maturation resulting in reduced plasticity (Lenneberg, 1967), significant improvements in perception can be made over relatively short periods of time. Training studies have demonstrated that participants can improve their perception of challenging foreign segmental (e.g. Logan, Lively & Pisoni, 1991) and suprasegmental distinctions (e.g. Wang et al., 1999; Wayland & Li, 2008), and that training can be generalized to new contexts and speakers. Furthermore, listeners were also capable of using non-native contrasts in a broader linguistic context, namely to distinguish lexical items (Curtin et al., 1998; Wong & Perrachione, 2007).

2.6 The present study

The current research utilized a seven-session lexical tone word training program for four groups of participants (tone and non-tone language background, subdivided into

musician versus non-musician). Participants were trained on the meanings of 15 novel vocabulary words, minimally distinguished by five Cantonese tones. A Cantonese tone identification task was employed before and after training to examine whether tone awareness before training predicts word learning proficiency and to determine if lexical training would transfer to an improvement on tone perception accuracy. Additionally, a musical aptitude task was administered to establish whether the musically-trained participants also possessed greater musical aptitude and more enhanced auditory acuity than the non-musicians.

2.6.1 Research Questions

As outlined in Sections 2.2.1.2 and 2.3.3, studies such as Wayland and Guion (2004) and Wong and Perrachione (2007) have investigated the separate effects of L1 tone language background and musical experience on non-native tone identification and word learning respectively. However, little research has been conducted on the interaction between these two factors, and what impact the combination will have on the acquisition of non-native tone words, as well as which factor facilitates the acquisition of new lexical items to a greater degree. This research sought to differentiate the relative influences of native language background and musical experience on second language tone word learning. This study investigated whether tone language experience or musical training facilitates the acquisition of new lexical items in Cantonese to any greater degree than the other at the initial stage of non-native tone learning. If no difference is found between either experiential factor, this might provide further support for an overlap in the cognitive mechanisms utilized for both language and music. On the other hand, if for instance, tone language speakers without musical training made larger gains in tone word

identification accuracy than non-tone language speakers with musical training, this might suggest that the relevant skills necessary for feature-to-word mapping are more domain specific. In other words, previous experience with utilizing tone to make lexical distinctions is more advantageous than musically-trained auditory acuity for success in mapping non-native featural information onto semantic content.

Additionally, the present study examined the relationship between musical ability and pitch discrimination skills and word-learning success, and whether such skills, as measured by self-reported musical experience, a musical aptitude task and a tone identification task, would be predictors of success in a foreign word-learning task. The current research sought to confirm the findings from Wong & Perrachione (2007), who reported that tone identification proficiency was a predictor of word learning success for English listeners, to another language group (Thai) to examine whether higher lexical tone identification accuracy predicts greater word-learning proficiency, that is, how linguistic experience might influence the feature-to-word learning transfer.

2.6.2 Hypotheses

The hypotheses for the present study are formulated with respect to three primary streams: 1) linguistic experience, 2) musical background, and 3) the interaction of these factors.

With regards to the first stream, best exemplified by a comparison of English and Thai non-musicians, the hypothesis is that the Thai group will attain greater proficiency at tone word identification than the English group. While Francis et al. (2008) reported that having a tone language background was not necessarily advantageous in identifying

lexical tones, this first hypothesis is motivated by the fact that training is a linguistic task, and Thai listeners will have more experience utilizing pitch distinctions to make lexical differentiations. Unlike the aforementioned studies, where participants needed just to identify the tone, the task in the present study involves mapping semantic content onto the tone word. Concerning the lexical tone identification task, given that previous findings have shown that linguistic experience can facilitate (e.g. Wayland & Guion, 2004) or inhibit tone perception (e.g. So, 2006), it is not certain whether Thais will have greater performance accuracy on Cantonese tone identification than the English group.

Next, based on studies such as Alexander et al. (2005) and Gottfried (2007) where musicians were found to be more accurate at identifying non-native tones, musically-experienced participants (both English and Thai) are expected to have higher levels of accuracy than their respective non-musician counterparts with regards to the pre- and post-training tone identification task. This pattern of performance is also hypothesized for the musical aptitude task, as musicians are anticipated to perform better than the non-musicians. We also hypothesize that, in line with results from Wong and Perrachione (2007), the musicians will be more successful during tone word training than the complementing non-musician groups, as their enhanced auditory perception will allow them to discern and retain the tonal contrasts with greater ease, allowing them to focus their attention on learning the appropriate meanings. Similar to Wong and Perrachione (2007), it is anticipated that success on both the tonal identification and musical aptitude tasks will predict lexical identification proficiency.

Thirdly, we hypothesize that the relative influence of the two factors in question (language background and musical experience) will produce a hierarchy of learning

success. The musically-trained participants with a tone language background (Thai musicians) are predicted to have the highest success in learning the Cantonese words, resulting from an additive effect of musical and linguistic tone experience. Next, the Thai participants without musical experience are expected to make larger gains during training than the English participants with musical experience. While English musicians have demonstrated superior pitch acuity (e.g. Alexander et al., 2005) and word learning proficiency (Wong & Perrachione, 2007) over non-musicians, the Thais may have an advantage given their previous experience using tone in a lexically significant manner. The performance of Thai non-musicians and English musicians will 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 second language word learning.

3: METHODOLOGY

3.1 Participants

A total of 77 adults participated in the present study, including 42 native Thai speakers and 35 native English speakers. A summary of their characteristics is provided in Table 3. Participants had 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. Additionally, they were all college-educated and possessed normal hearing and cognitive abilities. Both musical experience and aptitude were taken into consideration in constructing the definition of 'musician'. With regards to experience, musicians were defined as individuals who had undergone at least six years of continuous Western instrumental music training and had a current ability to play an instrument. Only musicians who received the majority of their musical training on Western instruments (e.g. piano, violin) were permitted to participate in order to avoid discrepancies in training styles between the Thai and English musician groups. 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, resulting in increased accuracy of musical pitch contour processing (Tervaniemi, Rytkönen, Schröger, Ilmoniemi, & Näätänen, 2001). In addition to self-reported experience, musicianship was also determined based on their musical aptitude scores (Gordon, 1989). Those scoring below the 20th percentile ranking were excluded. Non-musicians had less

than four years of musical experience, and no experience within the last five years. Participants scoring above the 80th percentile in their musical aptitude scores were excluded from the category. Participants were paid for their participation in this study.

The 34 Thai participants ultimately included in the study were students from Chulalongkorn University and Silpakorn University in Bangkok, Thailand. All participants were native speakers of the Bangkok dialect (Standard Thai), with English (ranging from poor to good proficiency) as their primary foreign language experience. Four participants possessed minimal experience with Japanese. Eighteen were non-musicians (10 male, 8 female; mean age: 22 years), with an average of 0.4 years of musical experience. Sixteen participants fit the criteria of musician (5 male, 11 female; mean age: 21 years). Their amount of musical experience ranged from 7 to 18 years, with a mean of 11 years. Fourteen of the musicians were music students in the Fine Arts Department at Silpakorn University. A detailed listing of the subjects' musical background is provided in Appendix A.

Native English participants were recruited from Simon Fraser University and the music department at the University of British Columbia. All 34 participants included were native speakers of English, some of whom possessed other foreign language experience, including French (20), German (5), Spanish (6), Japanese (3), Afrikaan (1), Tagalog (1), Hebrew (1), Dutch (1), Hungarian (1), Italian (1), and Finnish (1), but no tone language background. Sixteen participants were considered non-musicians (6 male, 10 female; mean age: 24), possessing an average of 1.5 years of musical experience. Eighteen participants fell under the established musicianship criteria (6 male, 12 female; mean age: 23). The amount of musical experience ranged from 9 to 22 years, with an

average of 16 years experience. Thirteen of the musicians were music students at the University of British Columbia.

Table 3 Characteristics of the four groups of participants

Group	Number (male, female)	Age (mean; range)	Mean years of musical experience	Musical aptitude percentile ranking
Thai musician	16 (5 M, 11 F)	21; 18-25	11	74.9
Thai non-musician	18 (10 M, 8 F)	22; 21-24	0.4	43.6
English musician	18 (6 M, 12 F)	23; 18-30	16	77.2
English non-musician	16 (6 M, 10 F)	24; 18-30	1.5	48.9

3.2 Stimuli

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

3.2.1 Training

Three Cantonese syllables with five Cantonese tones, for a total of 15 pseudo-words, were produced by four native Cantonese speakers (2 male, 2 female) for use in the training program. Unlike Wong and Perrachione (2007) who terminated training based on performance, participants in this study had a finite set of sessions to learn these words; thus, only 15 were used in order to enhance the learnability of the program. The stimuli were CV monosyllables (/k^{w}aj/, /tsou/, /wu/), comprised of a Cantonese initial consonant and a vowel common to Cantonese and the participants' native languages (Thai and

English). CV monosyllables were chosen, as the five selected Cantonese tones are only phonemically contrastive on open syllables. Closed syllables only possess three possible (level) tone patterns in Cantonese (Bauer & Benedict, 1997) and were thus excluded as a possible structure of stimuli. As these syllables were to be assigned meaning (in the form of a picture), an attempt was made to find CV syllables with common Thai/English phonemes that contained no semantic content in either language, so as to reduce lexical competition with existing words in participants' lexicons. As no such syllables appeared to exist, a compromise was made by inserting a Cantonese initial consonant followed by a vowel common to Cantonese, Thai and English (e.g. /k^{w}aj/). Stimuli including phonemes common to both Thai and English were included to facilitate learning, as studies have found that word learning can be inhibited when learners are faced with unfamiliar phonotactics (e.g. Ellis & Beaton, 1993a). As a result, pseudowords were necessitated in order to obtain these specific segmental and semantic specifications. 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). The Mid-Level (33) tone was not included, 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).

Each word was assigned a distinct meaning, as represented by a picture. A pictorial representation was chosen over a written translation because the participant groups had two different native language backgrounds, and the translations would not likely be identical cross-linguistically. Furthermore, the imageability of a concept has been shown to facilitate retrieval cues, and depicting word meanings as images may aid learning to a greater degree than providing its translated definition (Chun & Plass, 1996).

Specifically, concrete nouns were chosen as the meanings for the pseudowords, as they have been found to be easier to remember (Ellis & Beaton, 1993b). These pictures were selected from a set of 260 standardized pictures, controlled for visual complexity and cultural familiarity (Snodgrass & Vanderwart, 1980). Three categories of pictures were included for training (animals, human body parts, household items). Appendix B lists the five minimal quintuplets with their assigned meanings and pictures.

3.2.2 Pre-/post-training identification task

Two native Cantonese speakers (1 male, 1 female), different from the training speakers, produced a set of stimuli for the pre- and post-training tone identification task, which were distinct from the training syllables. Five Cantonese CV monosyllables containing common Thai/English phonemes were selected and produced with five Cantonese tones, for a total of 25 stimuli. The initial consonants were either voiced (/w/, /l/) or voiceless (/s/, /p/ /f/), to provide a variety of consonantal contexts. Five different vowels (/i/, /aj/, /ou/, /u/, /ej/) were included to enhance generalizability, so as to ensure that participants' identification performance holds across contexts, and not just for one particular vowel (see Appendix C for a full list of stimuli).

3.2.3 Familiarization

In order to acquaint the participants with the task procedures, a brief familiarization session was employed. This task utilized one syllable (/ji/), recorded by the female speaker from the pre-/post-training identification task. This syllable was similarly produced with five Cantonese tones.

3.3 Procedure

All participants completed both the pre-/post-training tasks as well as the training program 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. An overview of the procedure is provided in Table 4.

Table 4 Training program procedure break-down

Day 1	**Day 2**	**Day 3**	**Day 4**	**Day 5**
Musical aptitude task Pre-training identification task	Training session 1 Training session 2	Training session 3 Training session 4	Training session 5 Training session 6	Training session 7 Post-training identification task

3.3.1 Training

The participants engaged in a series of seven training sessions, approximately 30 minutes each, administered on four separate days over the course of two weeks. There was a minimum of one day and a maximum of 5 days between training days. The first three training days contained two training sessions per day, with approximately 15 minutes between sessions, and the last day contained one training session. Participants were trained to associate aurally-presented Cantonese words with visual representations of their meaning. The participants received training on all 15 words in every training session. As outlined in Figure 3, each session consisted of five blocks of three training words each, differing in tone and segmental information, followed by two review blocks with all 15 words and a session test. This format could be considered similar to the

"fading technique" (e.g. Jamieson & Morosan, 1986), whereby sessions are arranged with increasing levels of difficulty (in this case, moving from 3-word to 6-word to 15-word alternative choice blocks). Testing procedures, namely the three-word stimuli set, training block and session test format (illustrated in Figure 3), were modelled after training provided in (Wong & Perrachione, 2007).

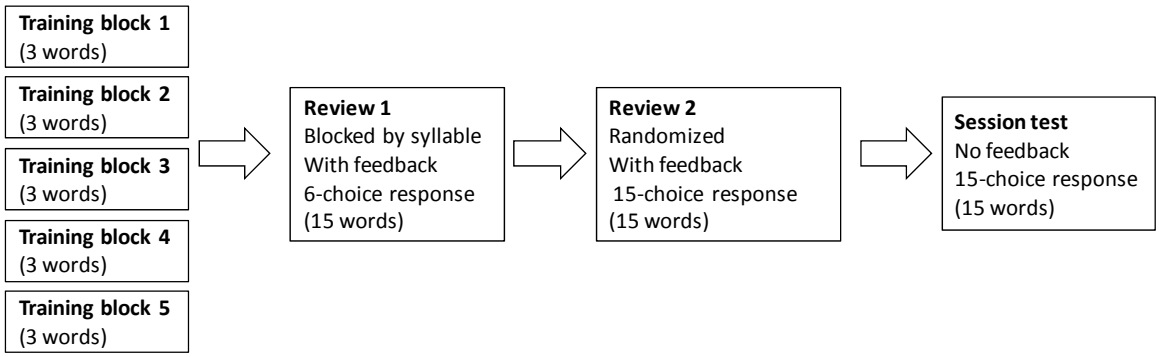


Figure 3 Training session setup. Training blocks contained 3 words differing in both tone and syllable. Review 1 was blocked by syllable in that stimuli were presented in 3 sequential sets of minimal tonal quintuplets.

The format of each training block (Figure 4) consisted of listening to four randomized repetitions (2 repetitions each from a male and a female speaker) of three words while viewing the visual representation of their meaning. This task was considered to simulate vocabulary learning, in that listeners would have to map a new lexical form onto a pictorially-represented meaning. Each block of words did not contain any 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 semantic clustering of new vocabulary items may actually be a detriment to learning (Finkbeiner & Nicol, 2003; Tamminen & Gaskell, 2008). Each block concluded with a small quiz on all three words

learned in that block. In these block quizzes, participants 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. There was no limit on response time, and they were also provided with feedback, informing them whether or not their answer was correct, displaying the correct answer and re-playing the aural stimulus. In total, each block consisted of 12 listening trials (2 speakers x 3 words x 2 repetitions) and 12 quiz trials (4 speakers x 3 words). It should be noted that listeners heard two speakers for the listening trials but were tested on four speakers to encourage generalization. Across sessions, the training blocks were counterbalanced such that each training block occurred as the initial block and final block at some point during the program. Each participant received the same ordering of training blocks.

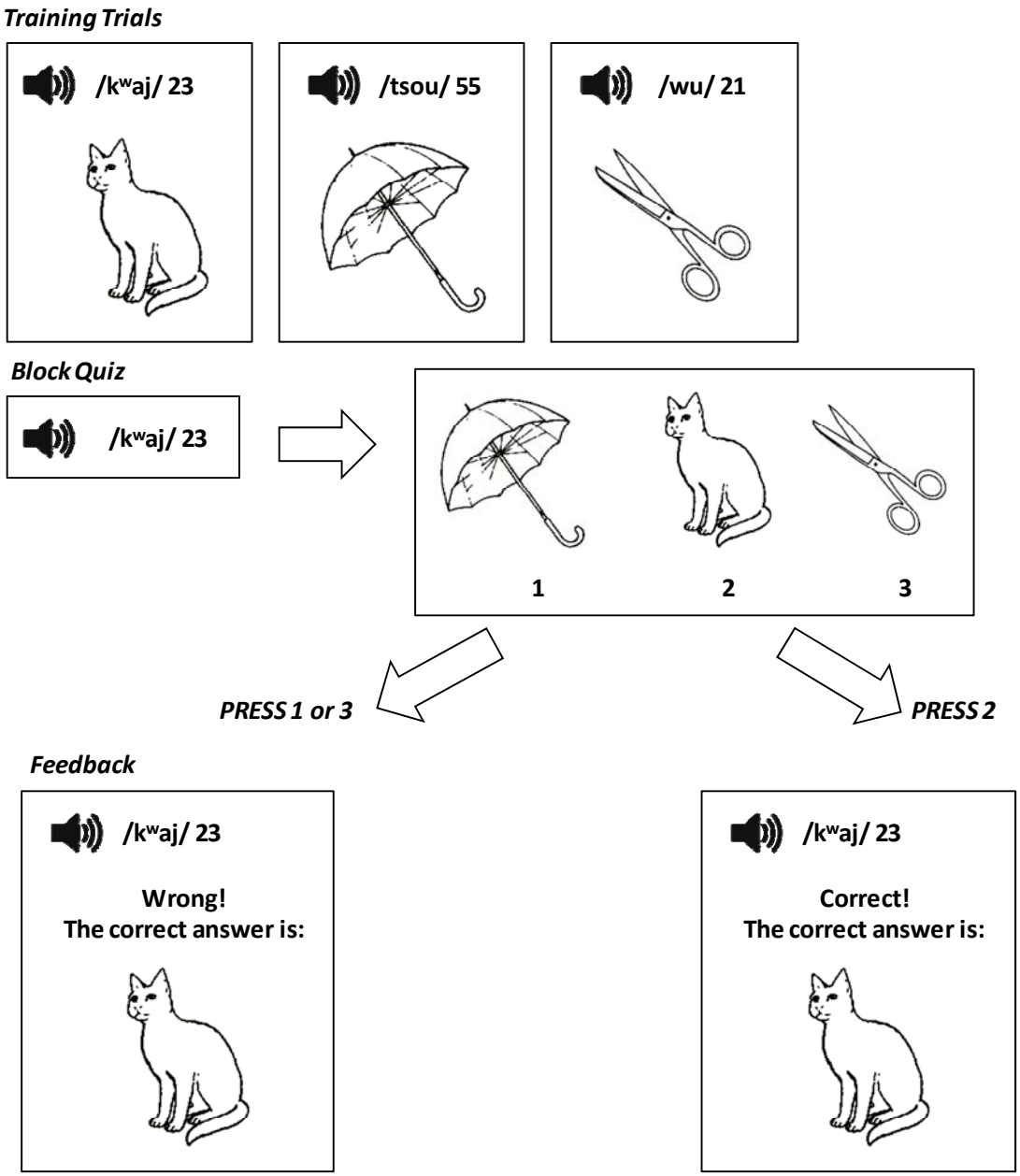


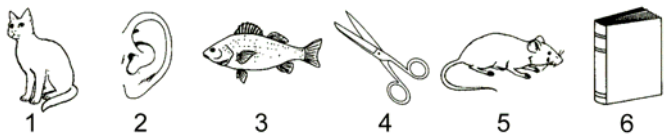
Figure 4 Block format. Note the IPA and tone numbers denote the auditory stimulus played for the listeners (they did not appear on the screen).

After completing the five training blocks, participants received two additional blocks reviewing the training items. Review 1 was comprised of all 15 words, blocked by syllable, produced by a female speaker from the training blocks. 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 (shown in Figure 5A), a minimal tonal quintuplet plus a foil image (a picture representing a segmentally-distinct word). Foils allowed us to see whether participants were still acquiring segmental information. Similar to the block quizzes, response time was not limited, and feedback was administered after each response, indicating the accuracy, displaying the correct picture and re-playing the stimulus item. In total, 15 trials comprised this review block (1 speaker x 15 words).

For Review 2, participants received two randomized repetitions of all 15 words from two training block speakers, choosing the meaning after each stimulus by selecting the appropriate picture from all 15 potential options on the screen, as illustrated in Figure 5B. Two repetitions were provided in this review to allow participants to become accustomed to choosing from all 15 choices, serving as a familiarization block for the session test. There was no limit on response time, and participants were given feedback on the correct answer and heard the stimulus again, similar to Review 1. In total, Review 2 contained 30 trials (2 speakers x 15 words).

A)



B)

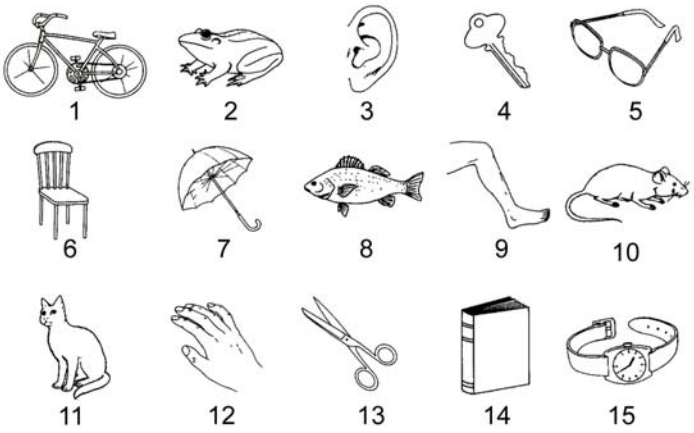


Figure 5 Response screens for the review blocks. A) is a sample six-choice response screen for Review 1 (for the syllable /k^{w}aj/). B) is the 15-choice response screen for Review 2.

At the end of each session, participants were tested on all 15 words learned in the training program without feedback. The session test followed the same format as Review 2, involving four randomized repetitions (1 production from each of the four speakers) of all 15 training words with an inter-stimulus-interval of 10 seconds. The participants heard a stimulus and chose the correct picture from a choice of 15 pictures. The session test was composed of 60 trials (15 words x 4 speakers). The results from this test were used to determine the progress of each participant.

3.3.2 Pre-/post-training identification task

All groups completed a lexical tone identification task, preceded by a brief familiarization section, before and after the training program. The familiarization portion allowed participants to become familiar with the five Cantonese tones and learn how to identify them. They first heard each Cantonese tone pronounced in isolation and viewed an appropriate 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 image, depicted in Figure 6. Feedback on the accuracy of their response as well as the correct answer was provided. 15 trials comprised this task (5 tones x 1 syllable x 1 speaker x 3 repetitions), which was considered sufficient time to be familiarized with the task and tones (Wayland & Li, 2008).

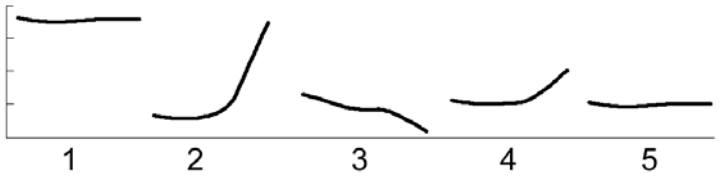


Figure 6 Response screen for pitch pattern identification task

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

3.3.3 Musical aptitude task

In addition to the familiarization and tone identification tasks, participants also completed a musical aptitude task before the tone identification task and commencing the training program. Gordon's *Musical Aptitude Profile* (Gordon, 1965) is an established standardized aptitude test designed to examine participants' proficiency at tonal imagery, rhythm imagery and musical sensitivity. However, at a running time of over three hours, it would be much too long to administer. For the purposes of this study, an abbreviated version, known as the Advanced Measures of Music Audiation (AMMA; Gordon, 1989),

was administered. This version is preferable, not only for its shorter duration (approximately 20 minutes), but because it was designed specifically for high-school students and adults. The tasks were designed such that no prior music training is required. The task involved listening to an audio MP3 and marking choices on a paper answer sheet. Participants listened to pairs of tonal melodies and indicated whether they were identical or different. If the latter, they specified whether it was a tonal or rhythmic variation. One pair of melodies would never contain both tonal and rhythmic discrepancies. Participants were provided three practice trials demonstrating the three possible melodic pairs (same, tonally different, rhythmically different). The AMMA yields tonal, rhythmic and composite scores.

4: RESULTS

This chapter describes the results of the analyses performed on the data from this study. Section 4.1 addresses the findings from the pre-/post-training identification task, including percent correct analyses (Section 4.1.1), confusion matrices (Section 4.1.2) and correlations with word learning attainment scores (Section 4.1.3). Section 4.2 outlines the results of the training sessions, beginning with overall improvement (Section 4.2.1), followed by training improvement trajectories (Section 4.2.2) as well as error type analyses (Section 4.2.3). Finally, the data from the musical aptitude task is provided in Section 4.3. Throughout this chapter, the significance level is set at 0.05; however, noteworthy results that are close to this significance level will also be reported and considered "marginally significant".

4.1 Pre-/post-training identification task

4.1.1 Percent correct analyses

Identification accuracy on these tasks was tabulated based on the proportion of correct responses by lexical tone (Figure 7). These mean percent correct scores were submitted to a mixed 3-way analysis of variance (ANOVA), with Test (pre, post) and Tone (55, 25, 21, 23, 22) as repeated measures, and Group (English non-musician, ENM; English musician, EM; Thai non-musician, TNM; Thai musician, TM) as a between-subjects factor.

A significant main effect of Test was obtained [$F(1,63)=49.787$, $p<0.0001$], indicating there was a significant overall improvement in accuracy from pre-training (52%) to post-training (65%) across groups.

Furthermore, the ANOVA also yielded a significant main effect of Tone [$F(4,63)=30.081$, $p<0.0001$]. Bonferroni-adjusted pairwise comparisons showed that High-Level tones (68%) had significantly higher identification accuracy rates than the High-Rising (51%, $p<0.0001$), Low-Rising (49%, $p<0.0001$) and the Low-Level (60%, $p=0.005$) tones averaged across tests and groups. Performance on the Low-Rising tones was also significantly worse than the Low-Falling and Low-Level tones ($p<0.0001$). Finally, the High-Rising tone also saw significantly poorer accuracy than Low-Falling and Low-Level tones across tests and groups ($p<0.0001$).

A significant main effect of Group was also obtained [$F(3,63)=28.232$, $p<0.0001$]. Post hoc (Tukey HSD) analysis found that EM had significantly higher accuracy rates overall (79%) than all other groups ($p<0.0001$). TM also performed significantly better (58%) than TNM (45%) across tests ($p=0.008$). No significant differences were found between TNM and ENM (51%, $p=0.330$), or between TM and ENM ($p=0.388$).

Finally, the interaction between Tone x Group was also significant [$F(12,63)=5.244$, $p<0.0001$], as was that of Test x Tone x Group [$F(12,63)=2.115$, $p=0.017$]. However, no significant interaction was found for Group x Test [$F(3,63)=1.037$, $p=0.383$].

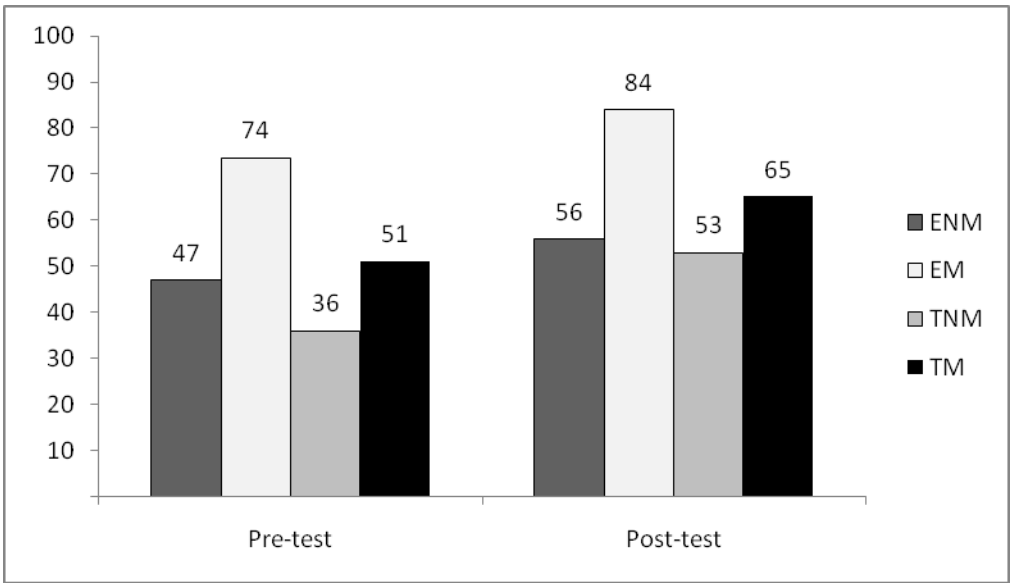


Figure 7 Mean percent correct for pre- and post-training identification tasks by group (E=English; T=Thai, NM=non-musician, M=musician)

4.1.1.1 Group differences by Tone

To examine group differences for which tones were easier or more challenging, 1-way ANOVAs for each Group with Tone as a repeated measure were conducted. Both English groups had the most success with identifying High-Level tones across tests ($p<0.002$). ENM also identified Low-Falling significantly better than Low-Rising overall ($p=0.005$). TNM's highest identification scores were for Low-Falling tones, significantly greater than Low and High-Rising and High-Level tones ($p<0.045$). Low-Level tones were also identified with greater accuracy than Low and High-Rising tones overall ($p<0.033$). For TM, Low-Falling, Low-Level and High-Level were their most successful tones, significantly more than Low-Rising ($p=0.049$). Low-Falling accuracy scores were marginally better than High-Rising scores as well ($p=0.093$).

4.1.1.2 Group differences by Test

2-way mixed ANOVAs were performed for each Test, with Tone as a repeated measures and Group as a between-subjects factor. For the pre-test, the main effect of Tone was significant [$F(4,63)=16.730$, $p<0.0001$], as was the Tone x Group interaction [$F(12,63)=5.369$, $p<0.0001$]. Subsequent 1-way ANOVAs with Group as the independent variable revealed highly significant differences for each tone on the pre-training task scores, illustrated in Figure 8. For the High-Level tone [$F(3,63)=34.635$, $p<0.0001$], post-hoc (Tukey HSD) analysis found TNM had significantly worse identification accuracy than all other groups ($p<0.010$), and that EM performed significantly better than all other groups ($p<0.0001$). No significant differences were found between ENM and TM ($p=0.180$). Post-hoc analysis of the 1-way ANOVAS for the High-Rising [$F(3,63)=11.601$, $p<0.0001$] and Low-Rising [$F(3,63)=7.228$, $p<0.0001$] tones showed that the only significant performance difference was between EM and all other groups ($p<0.006$), with EM obtaining significantly higher accuracy for these tones. A similar trend was found for the Low-Falling tone [$F(3,63)=20.930$, $p<0.0001$], with the EM performing significantly better than both ENM ($p<0.0001$) and TNM ($p=0.002$); however, no significant difference between EM and TM was obtained for this tone. Finally, both EM and TM significantly outperformed ENM ($p<0.043$) and TNM ($p<0.012$) on the Low-Level tone [$F(3,63)=10.735$, $p<0.0001$].

For the 2-way ANOVA on the post-test data, a significant main effect of Tone [$F(4,63)=24.481$, $p<0.0001$] and a significant Tone x Group interaction [$F(12,63)=2.933$, $p=0.001$] were also found. 1-way ANOVAs were performed on each tone for the post-training identification task data with Group as an independent variable, and highly significant group differences were yielded on all tones. Post-hoc (Tukey HSD) analysis

on the High-Level tone [$F(3,63)=11.688$, $p<0.0001$] revealed identical group patterns to the pre-test. EM had significantly higher performance than all other groups ($p<0.025$), and TNM performed significantly worse than the other groups ($p<0.031$). Although, the difference between TNM and TM was only marginally significant ($p=0.068$). Furthermore, consistent with the pre-test, post-hoc analysis for the High-Rising [$F(3,63)=12.849$, $p<0.0001$] tone found EM to be outperforming all other participant groups ($p<0.024$). TM also achieved significantly higher identification accuracy than TNM ($p=0.044$). ENM scores for this tone were not significantly different from TNM ($p=0.620$) and TM ($p=0.468$). The results of the post-hoc analysis for the Low-Falling tone [$F(3,63)=6.200$, $p=0.001$] indicated that EM obtained significantly better scores than ENM ($p=0.001$) and TNM ($p=0.043$), with no significant difference between EM and TM ($p=0.508$). TM also significantly outperformed ENM on this tone ($p=0.054$). Finally, post-hoc analyses for the Low-Rising [$F(3,63)=11.968$, $p<0.0001$] and Low-Level [$F(3,63)=6.112$, $p=0.001$] tones revealed that EM had a significantly larger proportion of correct responses than all other groups for both Low-Rising and Low-Level tones ($p<0.018$, $p<0.033$ respectively). No other significant group differences were found for these tones.

4.1.1.3 Pre- to post-training improvement for each Group

In order to determine whether each group made a significant improvement from pre- to post-test for all tones, 2-way ANOVAs were conducted for every group, with Test and Tone as repeated measures. For ENM, there was no main effect of Test [$F(1,15)=3.077$, $p=0.100$] or interaction between Test x Tone [$F(4,15)=.279$, $p=0.890$], indicating that there was no significant improvement in tone perception accuracy as a

result of lexical identification training. The results for EM demonstrated significant main effects of Test [$F(1,16)=16.302$, $p<0.0001$] and Tone [$F(4,16)=11.533$, $p<0.0001$], as well as a significant interaction of Test x Tone [$F(4,16)=2.664$, $p=0.040$]. Subsequent 1-way ANOVAs on each tone found significant increases in the proportion of correct responses after training for High-Rising ($p=0.023$), Low-Falling ($p<0.0001$), Low-Rising ($p=0.044$) and Low-Level ($p=0.025$) tones. The only tone that did not see a significant improvement for this group was the High-Level ($p=0.288$); although, this was likely due to a ceiling-effect, as EM achieved an average 94% accuracy on the pre-test for this tone. Similarly, results for TM saw significant main effects of Test [$F(1,14)=18.375$, $p<0.0001$] and Tone [$F(4,14)=5.283$, $p<0.0001$]. The interaction of Test x Tone was only marginally significant [$F(4,14)=2.261$, $p=0.074$], which may be attributed to virtually identical pre- and post-test scores (63.0% and 62.7%, respectively) for the Low-Level tone ($p=0.962$). However, performance on all of the other tones significantly improved after training for this group ($p<0.027$). Finally, the 2-way ANOVA for TNM data showed significant main effects of Test [$F(1,17)=28.461$, $p<0.0001$] and Tone [$F(4,17)=9.947$, $p<0.0001$], and a nearly significant interaction of Test x Tone [$F(4,17)=2.405$, $p=0.058$]. With the exception of the High-Rising tone ($p=0.084$), all of the tones saw significant increases in identification accuracy by the post-test ($p<0.003$).

In sum, these results indicate that all groups, with the exception of ENM, made significant improvements from pre- to post-test on tone identification accuracy. EM outperformed all other groups, and TM performed significantly better than TNM across tests and tones. Across tests, High-Level tones were the easiest for the English groups; whereas, Low-Falling and Low-Level tones were best for the Thai groups. On the pre-

test, EM outperformed most other groups on all tones, with TM performing better than the non-musicians on the Low-Level tone only. By the post-test, EM was still outperforming all other groups on High and Low Level tones and High and Low-Rising tones. TM was also significantly better than TNM and ENM on High-Rising and Low-Falling tones respectively.

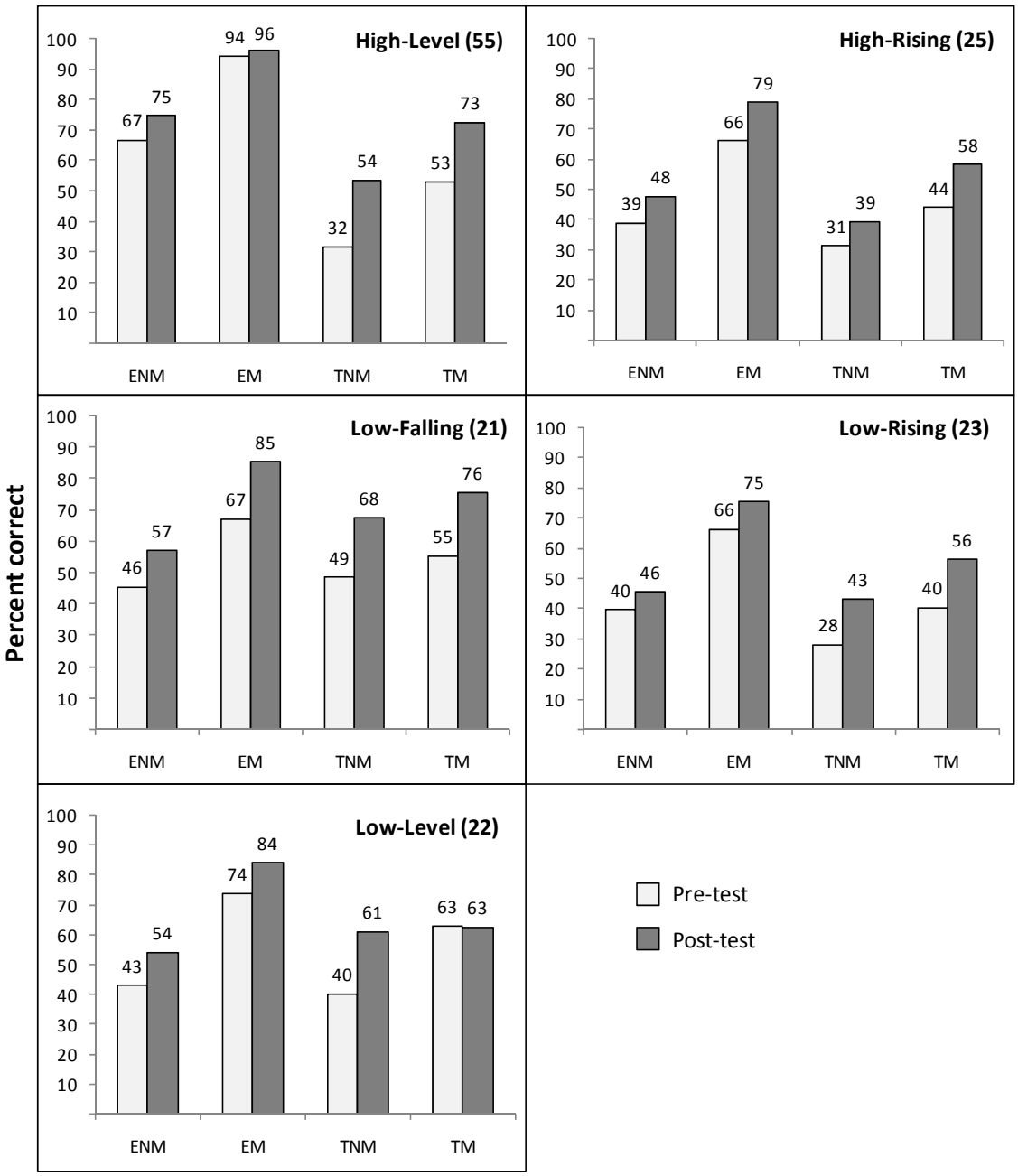


Figure 8 Mean percent correct for each tone by group for the pre-training and post-training identification tasks (E=English; T=Thai, NM=non-musician, M=musician).

4.1.2 Confusion analyses

Confusion matrices were constructed for the pre- and post-training identification data of each group (see Appendix D), in order to gain additional insight into how L1 experience influences lexical tone perception (e.g. Gandour, 1981; Francis et al., 2008). Constructing the confusion matrices involved creating 6 x 5 tables for each group and test, whereby the rows indicate the proportions of responses of a given tone (55, 25, 21, 23, 22, or no response), and the columns represent the tokens that they heard. For example, in Table 5, English musicians' responses for the High-Rising (25) tokens were High-Rising 66% of the time and Low-Rising (23) 32.5% of the time. Figure 9 provides the confusion patterns for each tone by group, averaged across pre- and post-tests.

Table 5 Sample confusion matrix for English musician pre-training ID scores

EM Pre-test					
Identified as	Token				
	55	25	21	23	22
55	94.2	0.0	0.6	2.5	13.6
25	0.0	66.1	0.3	16.4	0.0
21	0.0	0.6	67.2	2.5	7.2
23	0.8	32.5	2.5	66.4	4.2
22	4.4	0.3	29.4	11.4	73.9
No Response	0.6	0.6	0.0	0.8	1.1

4.1.2.1 High-Level (55)

A mixed 3-way ANOVA was performed with Group as a between-subjects factor, and Test (pre, post) and Confusion Pattern (55>25, 55>23, 55>22, 55>21)[1] as repeated

[1] It should be noted that the tone to the right of the ">" symbol indicates what the tone on the left is being misidentified as, here and elsewhere. For example, 55>23 indicates instances of High-Level tones being misidentified as Low-Rising tones. For the sake of brevity, tones in the confusion analysis section will be indicated using Chao's notation (1930): 55 (High-Level), 25 (High-Rising), 21 (Low-Falling), 23 (Low-Rising) and 22 (Low-Level).

measures. A significant main effect of Confusion Pattern was found [$F(3,63)=29.025$, $p<0.0001$], as well as a significant interaction between Confusion Pattern and Group [$F(9,63)=8.536$, $p<0.0001$]. This indicates a group difference in certain misidentification patterns across both tests. These patterns remained consistent across tests, as no significant Confusion Pattern x Group x Test interaction was obtained [$F(9,63)=0.883$, $p=0.542$].

1-way ANOVAs for each Confusion Pattern with Group as an independent variable found highly significant group differences for 55>25 and 55>23. TNM misidentified the High-Level tone as High-Rising significantly more than both English groups ($p<0.001$) and nearly TM ($p=0.058$). TNM also misidentified the High-Level as Low-Rising significantly more than the English groups ($p<0.002$) and TM ($p=0.031$). No significant group differences were found for the 55>22 confusion patterns ($p=0.117$).

Analysis of within-group tonal confusions revealed no significant difference in confusion pattern for ENM ($p>0.134$). EM misidentified this tone significantly more often as Low-Level (4%) than as High-Rising ($p=0.028$) or Low-Falling ($p=0.052$), and marginally more than Low-Rising ($p=0.088$). TNM demonstrated that 55>25 was the most common misidentification (29%), significantly more than 55>22 (6%, $p=0.001$) and 55>21 (3%, $p<0.0001$). Although, it was not significantly more common than 55>23 (16%, $p=0.126$). Similarly, TM's most common confusion patterns were 55>25 (17%), 55>25 (9%) and 55>22 (7.5%), which were also significantly more frequent than 55>21 ($p<0.008$).

4.1.2.2 High-Rising (25)

A 3-way ANOVA was carried out with Group as a between-subjects factor, and Test and Confusion Pattern (25>23, 25>55, 25>22, 25>21) as repeated measures. This yielded a significant main effect of Confusion Pattern [$F(3,60)=182.621$, $p<0.0001$]; although, no significant Group x Confusion Pattern interaction was obtained [$F(9,60)=0.998$, $p=0.444$]. There was a significant interaction between Test x Group x Confusion Pattern [$F(9,60)=5.382$, $p<0.0001$]. 2-way ANOVAs were performed for each group, with Test and Confusion Pattern as repeated measures. For ENM, no significant Test x Confusion Pattern interaction was found ($p=0.710$); however, a significant main effect of Confusion was obtained ($p<0.0001$). Pairwise comparisons revealed that 25>23 was the most common confusion pattern across tests ($p<0.0001$). 25>22 also occurred significantly more often than 25>55 and 25>21 ($p<0.050$). Analysis of the EM, TNM and TM results yielded similar results, with significant main effects of Confusion Pattern ($p<0.0001$), though Confusion Pattern x Test interactions were also significant ($p<0.011$) for all three groups. They confused 25 as 23 significantly more often than as 55, 22 or 21 ($p<0.009$) on the pre-test and the post-test ($p<0.014$). The 25>23 confusion pattern significantly dropped from pre- to post-test for TNM and EM ($p<0.015$).

4.1.2.3 Low-Falling (21)

A 3-way ANOVA with Group as a between-subjects factor, and Test and Confusion Pattern (21>22, 21>25, 21>23, 21>55) as repeated measures yielded a significant main effect of Confusion Pattern [$F(3,63)=75.990$, $p<0.0001$] and a significant Confusion x Group interaction [$F(9,63)=8.187$, $p<0.0001$]. There was also a significant Confusion Pattern x Group x Test [$F(9,63)=3.959$, $p<0.0001$]. A 2-way ANOVA for

ENM confusion scores with Test and Confusion Pattern as repeated measures revealed a significant main effect of Confusion Pattern ($p<0.0001$), but no significant interaction ($p=0.746$). Pairwise comparisons indicated that ENM had a significantly larger proportion of Low-Level misidentifications (31%) than High-Rising (5%, $p<0.0001$), Low-Rising (7%, $p<0.0001$) and High-Level (2%, $p<0.0001$) across tests. Similarly, EM consistently misidentified Low-Falling as Low-Level (29%), significantly more than any other confusion pattern ($p<0.0001$) in the pre-test. This was maintained into the post-test, as EM confused 21>22 (14%) with significantly greater frequency than the other tones ($p<0.047$). Indeed, both English groups confused 21>22 significantly more often than the Thai groups ($p<0.055$) on the pre-test (31% vs. 18%), and ENM continued to confuse it more frequently than the Thai groups on the post-test (31% vs. 10%, $p<0.001$).

TNM confused 21>22 (21%) significantly more often than 21>25 (7%) and 21>55 (6%, $p=0.002$ and 0.001, respectively) on the pre-test. No significant difference between the amount of 21>22 confusions and 21>23 (16%) confusions was found ($p=1.000$). After training, no significant difference was found between the proportion of 21>22 (11%), 21>23 (10%) and 21>25 (6%, $p>0.348$). 21>55 was confused significantly less often than 21>22 and 21>23 ($p<0.052$). For TM, a Test x Confusion Pattern interaction prompted further analyses, which indicated that on the pre-test, 21>22 (15%) and 21>23 (22%) confusion patterns were more frequent than 21>25 (3%, $p<0.028$) and 21>55 (2%, $p<0.002$), similar to TNM. By the post-test, there was also no significant difference between the frequency of 21>22, 21>23, and 21>25 confusion patterns ($p=0.381$). Only 21>55 was identified significantly less often ($p=0.030$). As compared to the English groups, the Thai listeners confused 21 as 23 significantly more frequently

before training ($p<0.014$). Though by the post-test, only TNM possessed a significantly greater proportion of 21>23 confusions than EM ($p=0.005$).

4.1.2.4 Low-Rising (23)

As noted in the previous section, participants across groups had the most difficulty identifying this tone. Similar analyses performed on this tone revealed a significant main effect of Confusion Pattern (23>25, 23>22, 23>55, 23>21) [$F(3,63)=23.747$, $p<0.0001$] and Confusion Pattern x Group interaction [$F(9,63)=4.178$, $p<0.0001$]. However, no Confusion Pattern x Group x Test interaction was obtained for this tone [$F(9,63)=1.248$, $p=0.268$]. 1-way ANOVAs for each confusion pattern with Group as an independent variable were carried out. No significant group differences were yielded for the 23>25 pattern ($p=0.980$), with an average 17% identifications of Low-Rising as High-Rising across groups. ENM did confuse 23>22 (25%) significantly more than TNM (16%) and EM (8%, $p<0.054$), with the difference between ENM and TM reaching marginal significance (16%, $p=0.065$). TNM also significantly misidentified 23 as 55 (15%) more often than all other groups ($p<0.011$). The Thai groups also confused 23>21 (10%) significantly more frequently than EM (2%, $p<0.026$), but not more so than ENM (8%).

To examine which tones were confused as Low-Rising within each group, a series of 1-way repeated measures ANOVAs were performed. The English groups had similar patterns, in that 23>25 and 23>22 occurred significantly more often than the other confusion patterns overall ($p<0.015$). TNM exhibited a more even distribution of confusion patterns, with no significant differences found between then ($p=0.747$).

However, TM did see a significant difference between confusions, with 23>25 and 23>22 occurring more frequently than the others, though only marginally (p>0.066).

4.1.2.5 Low-Level (22)

A 3-way ANOVA with Test and Confusion Pattern (22>55, 22>23, 22>21) as repeated measures and Group as a between subjects factor yielded a significant main effect of Confusion Pattern [$F(3,63)=25.330$, $p<0.0001$]. However, no significant Confusion Pattern x Group or Confusion Pattern x Group x Test interactions were found [$F(9,63)=1.619$, $p=0.112$; $F(9,63)=0.591$, $p=0.803$]. Pairwise comparisons indicated that listeners across groups misidentified 22 as 55 (15%) significantly more than any other tone, though the difference was only marginal between 22>55 and 22>23 (10%, $p=0.071$). No significant difference between the proportion of 22>23 and 22>21 confusions (9%, $p=1.000$) was found.

To summarize, High-Level tones were most commonly misidentified as High-Rising by Thai groups, but as Low-Level by EM. For High-Rising tones, 25>23 was the most common misidentification for all groups. Low-Falling tones were wrongly identified as Low-Level most frequently by English groups; however, Thai groups saw a greater split in their confusion patterns (21>22, 21>23 and 21>25). For the Low-Rising tones, 23>25 and 23>22 were the most common confusion patterns for the English groups and TM; although, TNM had a fairly even distribution of misidentifications (23>25, 23>22, 23>55, 23>21). Finally, Low-Level was most frequently confused for

High-Level across groups.

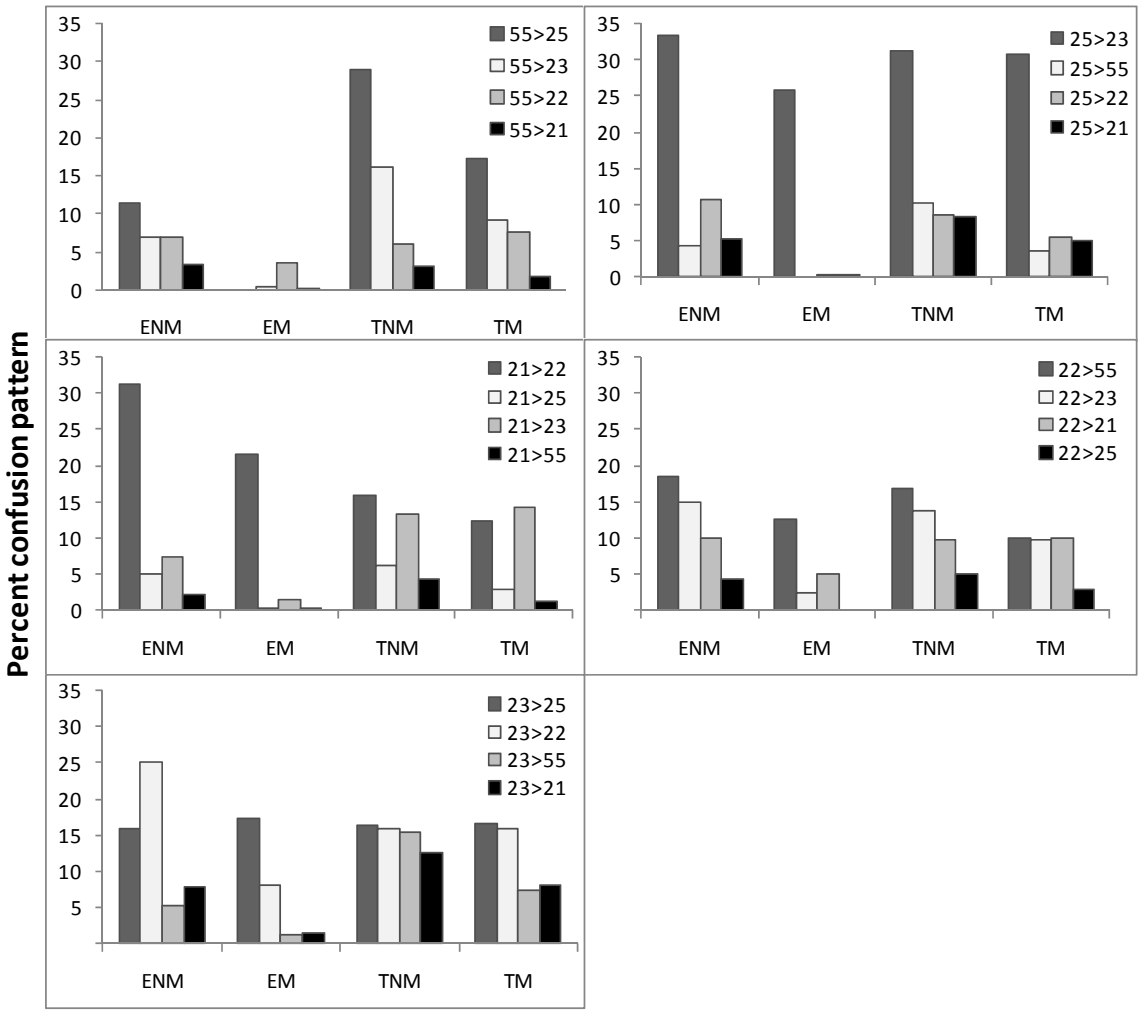


Figure 9 Mean percent confusion patterns (averaged across pre- and post-tests) by group.

4.1.3 Pre-test identification and Overall Attainment

Previous work reported that pitch pattern identification accuracy significantly predicted word learning proficiency (Wong & Perrachione, 2007). In order to examine whether such findings would hold for the present study, a linear regression model was employed on the data from all groups, with training Session 7 percent correct scores as the dependent variable and pre-training identification scores as the predictor. Mean percent correct scores for the pre-training identification task were 36% (TNM, s.d.=10),

47% (ENM, s.d.=15), 51% (TM, s.d.=14) and 74% (EM, s.d.=8). Results, as illustrated in Figure 10, indicate that pre-test scores were a significant predictor of lexical identification by the end of training (R^2=0.207, F(1,65)=16.948, p<0.0001). Linear regression models were constructed for each group individually, with Session 7 scores as the dependent variable and pre-training scores as the predictor. Results revealed that pre-test scores significantly predicted tone word learning success for most groups, though only marginally so for TNM (ENM, R^2=0.532, F(1,14)=15.944, p=0.001; TM, R^2=0.338, F(1,13)=6.629, p=0.023; TNM, R^2=0.197, F(1,16)=3.922, p=0.065). However, for EM, the pre-test and Session 7 scores were not significantly correlated (p=0.466).

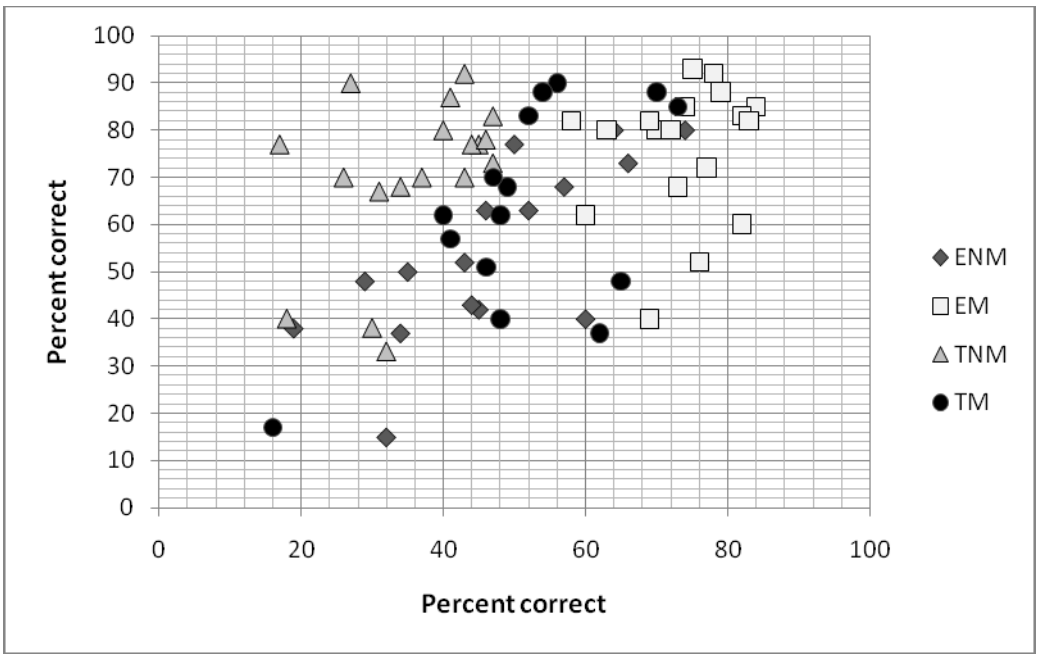


Figure 10 Mean percent correct pre-training identification task scores (X-axis) against mean percent correct Session 7 scores (Y-axis) by group.

4.1.4 Summary

The lexical tone identification task examined listeners' abilities to identify non-native tones before and after a training program whose primary focus was on improving

tone word identification accuracy. With regards to overall accuracy levels across tests and tones, EM significantly outperformed all other groups. TM also was significantly more accurate than TNM. All groups, with the exception of ENM, significantly increased their tone identification accuracy after training. Specifically, EM saw an improvement on all tones except High-Level. TM improved on every other tone except Low-Level, and TNM made accuracy gains on all tones except High-Rising.

On the pre-training task, EM had significantly higher identification accuracy rates than all other groups on High-Level, High and Low-Rising tones. They also were significantly more accurate than ENM and TNM on Low-Falling and Low-Level tones. TM also performed significantly better than ENM and TNM on Low-Level tones. By the post-test, EM maintained their higher accuracy scores on High-Level, High and Low-Rising as well as Low-Level tones as compared to all other groups. TM was also better than TNM on the High-Rising tone. For Low-Falling tones, EM and TM were significantly more accurate than ENM. EM was also significantly more accurate on these tones than TNM.

Finally, tonal confusion patterns indicated that High-Level tones were misidentified as High-Rising significantly more often by Thai groups than the English groups across tests. It was also confused as Low-Level by EM more often than other confusion patterns. High-Rising tones were largely misidentified as Low-Rising for most groups on both pre and post tasks. Low-Falling tones were confused most frequently as Low-Level by the English groups; the Thais confused them as either Low-Level, Low-Rising or High-Rising. Low-Rising tones were largely confused with High-Rising or

Low-Level tones across groups. Lastly, Low-Level was most often misidentified as High-Level for most groups.

4.2 Training

4.2.1 Overall Improvement & Attainment

As outlined in Section 3.3.1, tone word identification performance was assessed from a session test concluding each training session. To evaluate overall improvement over the course of training, mean percent correct for each tone was calculated for the first and last training sessions. These scores were input into a 3-way mixed-design analysis of variance (ANOVA) with Session (1, 7) and Tone (55, 25, 21, 23, 22) as repeated measures, and Group (ENM, EM, TNM, TM) as a between-subjects factor. The mean percent correct scores by group, averaged across tones, are depicted in Figure 11.

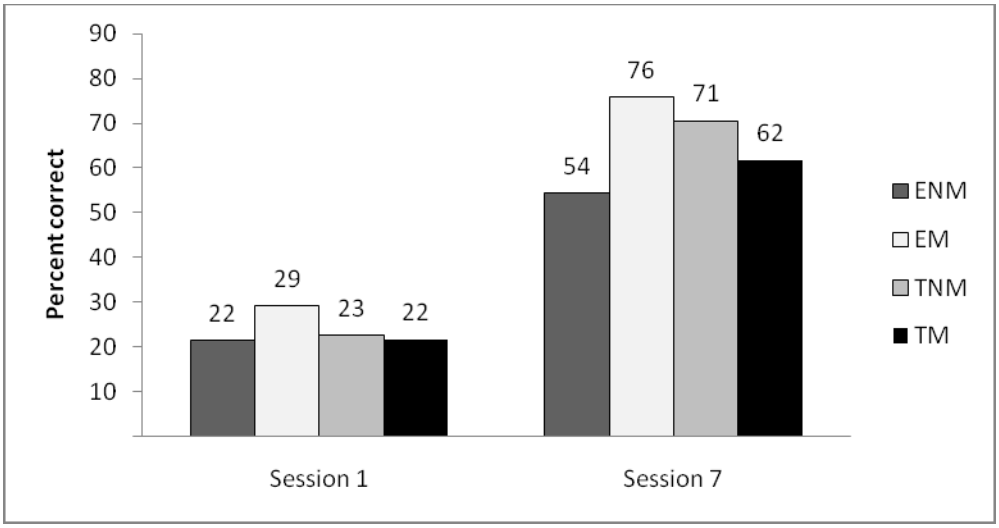


Figure 11 Mean percent correct scores (across tones) for each group for training Session 1 and Session 7

A significant main effect of Session was obtained [$F(1,16)=504.402$, $p<0.0001$], indicating that listeners were able to significantly improve their tone word identification

accuracy scores after training (*M*=24% to 66%). The ANOVA also yielded a significant main effect of Group [F(3,64)=4.192, *p*=0.009]. Post-hoc (Tukey HSD) analysis indicated that EM significantly outperformed ENM across the first and last sessions (*p*=0.008). The 3-way ANOVA also found a significant main effect of Tone [F(4,64)=37.366, *p*<0.0001], indicating that lexical items with certain tones were significantly easier than others across groups and sessions.

4.2.1.1 Group differences by Session

Furthermore, a significant Session x Group interaction was also found [F(3,64)=3.420, *p*=0.022]. To further examine whether each group significantly improved from the first to last session, 1-way ANOVAs for each group were performed, averaging across tones, with Session as the independent variable, confirming that all groups made a highly significant improvement as a result of training (*p*<0.0001), with a mean increase of 42%.

Subsequent 1-way ANOVAs for each session with Group as the independent variable were conducted. No significant group differences were found for the first session [F(3,64)=2.031, *p*=0.118]; however, significance was achieved for the last session [F(3,64)=4.689, *p*=0.005]. Post-hoc (Tukey HSD) analysis revealed that by the last session, EM (76%, s.d.=14) and TNM (71%, s.d.=17) had significantly higher accuracy rates than ENM (54%, s.d.=19; *p*<0.0001, *p*=0.053 respectively). There were no significant differences found between TM (62%, s.d.=22) and any of the other groups (*p*>0.112), or between EM and TNM (*p*=0.816).

4.2.1.2 Group differences by Tone

In addition, a significant Tone x Group interaction was also found [F(12,64)=1.926, *p*=0.032]. However, no significant interaction between Session x Tone x Group was obtained [F(12,64)=1.476, *p*=0.133]. Thus, 1-way ANOVAs for each group, averaging across both sessions, with Tone as the independent variable were performed. Results showed a significant difference by tone for all groups (ENM [F(4,15)=21.018, *p*<0.0001], EM [F(4,17)=9.799, *p*<0.0001], TNM [F(4,17)=6.571, *p*=.004], TM [F(4,15)=8.521, *p*<0.0001]). Figure 12 provides the mean percent correct for each tone (averaged across session 1 and 7) by group. Pairwise comparisons with Bonferroni-adjustment were conducted on the data for each group. For EM, Low-Rising tone lexical items were identified significantly less accurately than the High-Level (*p*<0.0001), Low-Falling (*p*=0.001) and Low-Level (*p*=0.036) ones. Additionally, items with a High-Level tone possessed a significantly higher identification accuracy score than Low-Level tone items (*p*=0.044). ENM performed significantly better on High-Level tone items than all of the other tone words across sessions (*p*<0.02). Similarly, TM found High-Level tone lexical items to be significantly easier to identify than High-Rising (*p*=0.003), Low-Rising (*p*=0.001) and Low-Level (*p*=0.017) words. Finally, the only significant difference in identification accuracy for TNM was between Low-Rising tone words and High-Level (*p*=0.001) ones, with Low-Falling (*p*=0.058) items nearing significance.

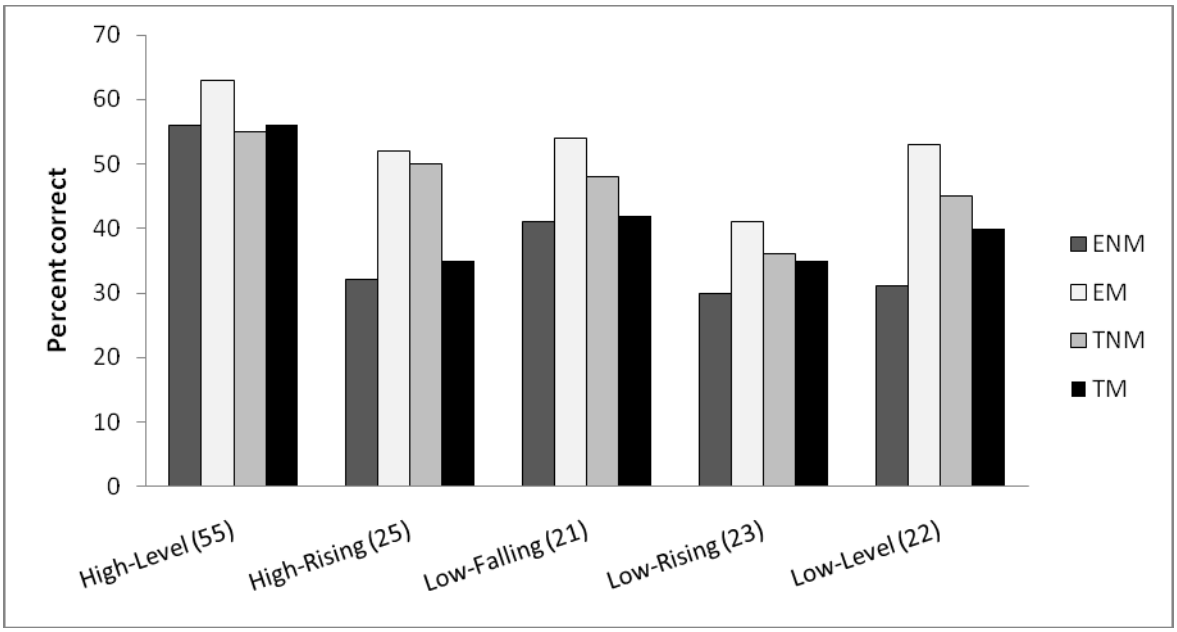


Figure 12 Mean percent correct (averaged across session 1 and 7) for each tone by group

In order to examine any group differences in tone word identification accuracy by tone, 1-way ANOVAs were performed for each tone with Group as the independent variable. No group differences in accuracy were found for lexical items with High-Level (p=0.266), Low-Falling (p=0.143) or Low-Rising (p=0.168) tones, when averaged across both sessions. However, for High-Rising tone words, a significant group difference was obtained [F(3,64)=6.451, p=0.001]. Post-hoc (Tukey HSD) analysis showed that ENM were significantly less accurate at identifying these tone words than EM (p=0.004) and TNM (p=0.013). Furthermore, TM also had significantly lower performance than EM on these lexical items (p=0.020). Finally, the 1-way ANOVA for the Low-Level tone words yielded significant group differences [F(3,64)=4.494, p=0.006], with EM's performance (53%) being significantly better than ENM's performance (31%) for these words.

Thus, with regards to overall improvement, all groups saw a significant increase in tone word identification accuracy from the first to last session. No significant group

differences were found at the beginning of training; however, by the final session, EM and TNM performed significantly better than ENM. High-Level tone word items were largely the easiest to identify for most groups, and Low-Rising tone words were often the most challenging. Across the first and last sessions, EM identified High-Rising and Low-Level tone words significantly more accurately than ENM. TNM also had higher accuracy on High-Rising words than ENM.

4.2.2 Training Improvement Trajectories

As most training days involved two sessions (with the exception of the last day), averages of each day's training sessions were compiled in order to better examine how performance varied over the course of the training program. Accordingly, results will be reported with respect to Day (1-4). Mean percent correct scores by Day for each group were tabulated (Figure 13) and submitted to a mixed 3-way ANOVA, with Day (1-4) and Tone (55, 25, 21, 23, 22) as repeated measures, and Group (ENM, EM, TNM, TM) as a between-subjects factor.

4.2.2.1 Group differences by Day

A significant main effect of Day was obtained [$F(3,62)=265.079$, $p<0.0001$]. Bonferroni-adjusted pairwise comparisons revealed that listener groups saw significant increases in performance between each day ($p<0.0001$), with mean identification accuracy scores moving from 31% on Day 1 to 51% on Day 2, and 61% on Day 3 to 65% on the final day, indicating that identification accuracy significantly improved over the course of training across groups

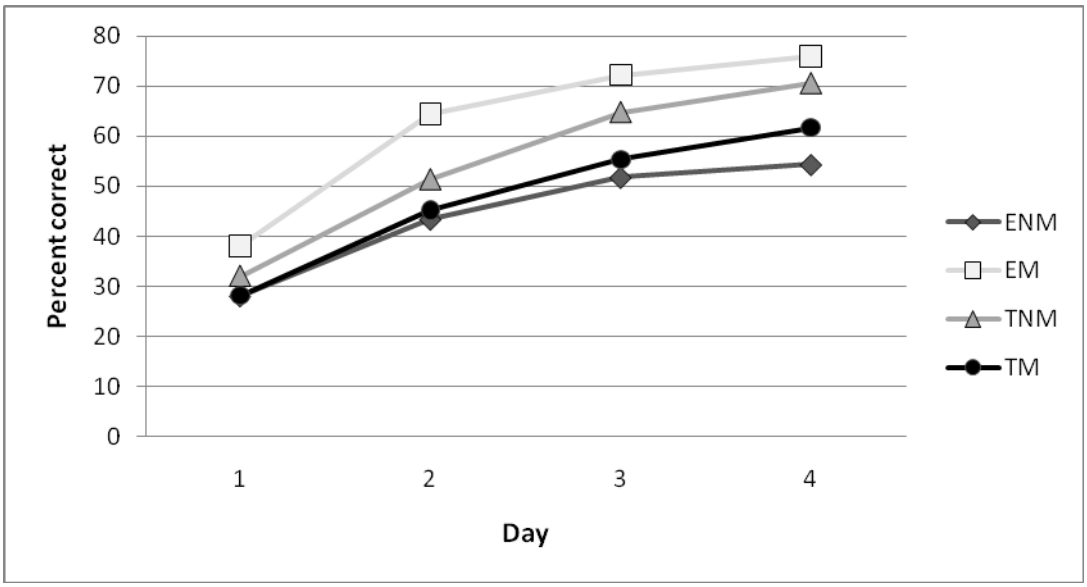


Figure 13 Mean percent correct by day for each group

The 3-way ANOVA also yielded a significant Group x Day interaction [F(9,62)=2.529, *p*=0.009], indicating a group discrepancy in performance on certain training days. In order to examine any group differences in performance for each day, a series of 1-way ANOVAs for each Day with Group as the independent variable were performed. Consistent with the findings presented in Section 4.1.1, there were no significant group differences in performance on Day 1 [F(3,63)=2.281, *p*=0.088]. However, a significant group discrepancy emerged by Day 2 [F(3,63)=4.771, *p*=0.005], and remained on Day 3 [F(3,63)=5.169, *p*=0.003] and Day 4 [F(3,63)=4.689, *p*=0.005]. Post hoc (Tukey HSD) analysis indicated that by Day 2, EM's proportion of correct responses (65%, s.d.=16) was significantly higher than ENM (43%, s.d.=17; *p*=0.007) and TM (45%, s.d.=20; *p*=0.015). TNM identification accuracy (51%, s.d.=21) did not differ significantly from any of the groups (*p*>0.146). This pattern of group differences was maintained on Day 3, whereby EM (72%, s.d.=17) had significantly better

performance than ENM (52%, s.d.=16; p=0.004) and TM (55%, s.d.=19; p=0.025). Similarly, TNM's accuracy rate (65%, s.d.=16) was not significantly different from the other groups (p>0.120). By the final day, we see a shift in performance patterns, as both EM (76%, s.d.=14; p=0.005) and TNM (71%, s.d.=17; p=0.053) had higher identification accuracy scores than ENM (54%, s.d.=19). No significant differences between TM (62%, s.d.=22) and the other groups were found (p>0.112).

To examine each group's specific improvement pattern, multiple 1-way ANOVAs by Group with Day as the independent variable were conducted. All groups saw significant differences in their identification performance (p<0.0001), and the gains made by each group were similar. For each group's 1-way ANOVA, Bonferroni-adjusted pairwise comparisons indicated that all groups saw a significant increase in accuracy from Day 1 to Day 2 (p<0.0001) and from Day 2 to Day 3 (p<0.006). However, no significant gains in identification accuracy were made between Days 3 and 4 for any group (p>0.084).

4.2.2.2 Patterns of training by Tone

In addition, the 3-way ANOVA yielded a significant main effect of Tone [F(4,62)=67.976, p<0.0001], as well as a significant Day x Tone interaction [F(12,62)=4.621, p<0.0001]; however no significant Day x Tone x Group interaction was found [F(36,62)=1.300, p=0.115]. Subsequent 1-way ANOVAs on each Tone with Day as a factor revealed that all groups saw significant increases in accuracy over the course of the four days (p<0.0001). Across groups, there were significant gains in accuracy for lexical items with High-Level, Low-Falling and Low-Level tones between Days 1 and 2 and Days 2 and 3 (p<0.0001); although, they failed to make a significant improvement

from Day 3 to 4 for these tones ($p>0.147$). For High-Rising and Low-Rising items, however, listener groups saw a significant performance improvement for each successive day ($p<0.032$).

4.2.2.3 Degree of Improvement

As indicated in Section 4.2.2.1, all groups made significant improvements in performance over the first three days of training; however, this does not reveal the degree of improvement made over the course of training and among groups, and whether this remained consistent over training. To examine any group differences in the degree of improvement over training, three improvement amounts were calculated by subtracting mean percent correct scores of Day 1 from Day 2, of Day 2 from Day 3 and of Day 3 from Day 4 (Figure 14). A 2-way mixed ANOVA with Rate change (Day 1-2, Day 2-3, Day 3-4) as repeated measures and Group (ENM, EM, TNM TM) was computed, which yielded a significant main effect of Rate change [$F(2,62)=46.838$, $p<0.0001$]. Pairwise comparisons with Bonferroni-adjustment showed a significant difference between each rate of change ($p<0.002$), indicating that the amount of improvement made from Day 1 to 2 (20%) was significantly more than from Day 2 to 3 (10%) and from Day 3 to 4 (4%) across groups.

A significant Rate change x Group interaction was also found [$F(6,62)=2.178$, $p=0.049$]. A 1-way ANOVA for each Rate change with Group as the independent variable found there was only a significant group difference in the amount of improvement made for Day 1 to 2. Post-hoc analysis indicated that EM had a significantly higher degree of improvement than ENM (26% vs. 15%, respectively). Both EM and ENM made significant improvements in their accuracy from Day 1 to 2;

however, these results illustrate that EM made a larger amount of improvement than ENM.

Furthermore, 1-way ANOVAs for each Group with Rate change as an independent variable revealed their learning curves over the course of training. All groups saw a significant difference in Rate change ($p<0.005$), suggesting that within each group, there was a discrepancy in the degree of improvement between training days. EM saw a more rapid decline in the amount improved from Day 1 to 2 as compared to the difference between Day 2 to 3 ($p<0.0001$) and Day 3 to 4 ($p<0.0001$). However, no significant difference was found between the percent different of Day 2 to 3 and Day 3 to 4 ($p=0.312$). The other three groups shared similar patterns, but demonstrated a more gradual decline in improvement difference, with no significant difference between Day 1 to 2 and Day 2 and 3 amounts ($p>0.184$), as well as between Day 2 to 3 and Day 3 to 4 amounts ($p>0.081$). The only significant difference in Rate Change for these three groups was found for the degree of improvement from Day 1 to 2 and Days 3 and 4 ($p<0.005$).

In sum, no significant group differences in performance accuracy were found on the first day; however, EM outperformed ENM and TM on Day 2 and 3. By the last day, TNM and EM were both outperforming ENM. The degree of improvement made by all groups significantly decreased with each successive day in the training program. EM had a significantly greater degree of improvement from Day 1 to 2 than ENM. EM also had a significantly larger initial decrease in improvement gains than other groups, who showed a more gradual improvement decline over the course of training .

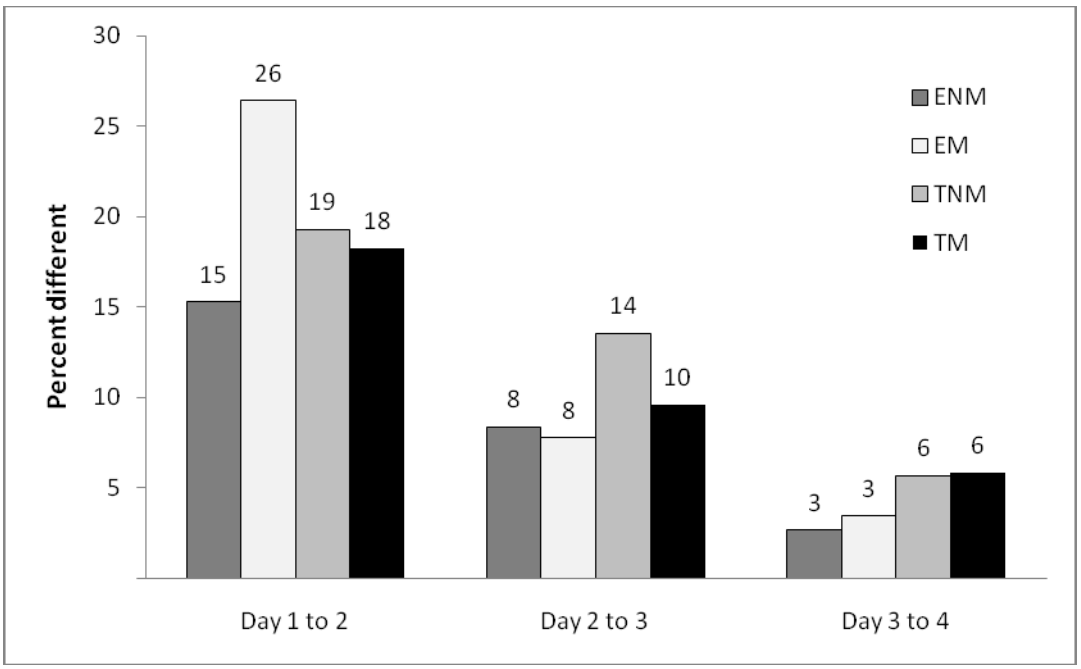


Figure 14 Mean difference in percent correct from Day 1 to 2, Day 2 to 3 and Day 3 to 4 by group

4.2.3 Error type analysis

In order to successfully acquire the lexical items presented in this task, participants needed to learn both the segmental and tonal components for each item. Consequently, identification errors could result from misidentifying either the segmental or tonal components, or both. Therefore, the type of errors were calculated for the first and last sessions by determining what proportion of the overall percentage of errors was segmental, tonal or both. An error was designated tonal if the participant's response matched the correct segmental information but was an incorrect match for the tonal pattern (e.g. an answer of *wu* [Low-Rising] when the stimulus was *wu* [High-Rising]). Segmental errors were determined based on whether the participant responded with the correct tonal pattern but the wrong syllable (e.g. answering $k^{w}aj$ [High-Rising] for the stimulus *tsou* [High-Rising]). Finally, participants could also misidentify both the tonal

pattern and the segmental information (e.g. answering *wu* [High-Rising] for the stimulus *tsou* [Low-Rising]), which was classified as a both (tone/segment) error.

The percentages were submitted to a mixed 3-way ANOVA, with Session (1, 7) and Error Type (segmental, tonal, both) as repeated measures and Group as a between-subjects factor. The main effect of Error Type was highly significant [$F(2,64)=1003.569$, $p<0.0001$], and Bonferroni-adjusted pairwise comparisons indicated that the proportion of tonal errors (78%) was significantly greater than segmental (6%, $p<0.0001$) or both errors (14%, $p<0.0001$), across groups and the first and last sessions. There was also a significant interaction of Error Type x Session [$F(1,64)=8.937$, $p=0.004$]. Further analysis indicated that in the first session, 63% of the errors made were tonal, which was significantly more than both (26%, $p<0.0001$) and in turn significantly more than the segmental (8%, $p<0.0001$). By the last training session, 93% of the errors being made were tonal, significantly more than segmental (3%, $p<0.0001$) and both (3%, $p<0.0001$). Across groups, the proportion of segmental and both errors significantly dropped from the first to last session ($p<0.0001$), while the proportion of errors that were tonal significantly increased ($p<0.0001$).

Finally, the 3-way ANOVA yielded a significant Error Type x Group interaction [$F(2,64)=192.633$, $p<0.0001$]. However, no significant Error Type x Session x Group interaction was found [$F(6,64)=1.773$, $p=0.110$]. 1-way ANOVAs for each error type found a significant group difference for both segmental [$F(3,64)=8.663$, $p<0.0001$] and tonal errors [$F(3,64)=3.160$, $p=0.031$], across sessions. Post-hoc (Tukey HSD) analyses indicated that EM had significantly more segmental errors overall than all other groups

($p<0.009$), and that TNM had a significantly higher proportion of tonal errors compared to EM ($p=0.040$) and nearly TM ($p=0.070$).

4.2.4 Summary

With regards to overall attainment (comparing first and last training sessions), lexical identification accuracy significantly improved for each tone item set across groups. All groups started training with similar accuracy rates and significantly increased their overall accuracy after training. Group differences arose, however, by the final training session, in that EM and TNM achieved higher overall attainment levels than ENM. TM did not differ significantly from any of the groups. Furthermore, High-Level tone words were consistently the most accurately identified, and Low-Rising tone words were the most challenging across groups. Overall, ENM had significantly worse performance than EM on Low-Level tone words, and lower accuracy rates than both EM and TNM on High-Rising tone lexical items.

Results of the training pattern analyses showed that accuracy improvements were made on each successive training day across groups. By Day 2, group differences emerged in identification accuracy, with EM performing significantly better than TM and ENM, which was maintained on Day 3. TNM and EM both achieved higher accuracy rates than ENM on Day 4, with no significant differences for TM. Accuracy gains were made on all tone item sets from Day 1 to 2 and from Day 2 to 3; however, only High-Rising and Low-Rising items also saw accuracy improvements from Day 3 to 4 across groups. For the degree of improvement, all groups made significantly smaller gains on each successive training day. The only group difference arose between Days 1 and 2, where EM made a significantly larger gain in accuracy than ENM.

Finally, error-type analysis revealed that the majority of tone word identification errors made in the first session were tonal, and that this was maintained in the last session. Across the first and last sessions, EM had significantly more segmental errors than all other groups, and TNM had more tonal errors as compared to EM.

4.3 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. A 1-way ANOVA with ranking scores as the dependent variable and Group as the independent variable revealed a significant group difference [$F(3,63)=23.924$, $p<0.0001$]. Post-hoc (Tukey HSD) analysis indicated that both musician groups (EM, TM) achieved significantly percentile rankings than both the non-musician groups ($p<0.0001$).

These rankings were then plotted against the percent correct identification for Session 7 (Figure 15). Results point to a positive correlation between musical aptitude percentile rankings and overall attainment level ($r= +0.242$, $p=0.046$). This suggests that across groups, participants with higher musical aptitude scores tended to achieve higher tone word identification proficiency by the final session. However, when broken down by group, correlation analyses revealed that this correlation was only significant for the English groups. Linear regression models were constructed for each group with musical aptitude percentile ranking as the predictor and Session 7 scores as the dependent variable. For both EM and ENM, musical aptitude ranking was a significant predictor of success in the tone word learning task (EM: $R^2=0.305$, $F(1,16)=7.028$, $p=0.017$; ENM: $R^2=0.288$, $F(1,14)=5.651$, $p=0.032$). Aptitude scores were not significantly correlated with Session 7 percent correct scores for either Thai group ($p>0.696$). These results

indicate that performance variation for overall attainment within groups can in part be explained by musical aptitude percentile rankings.

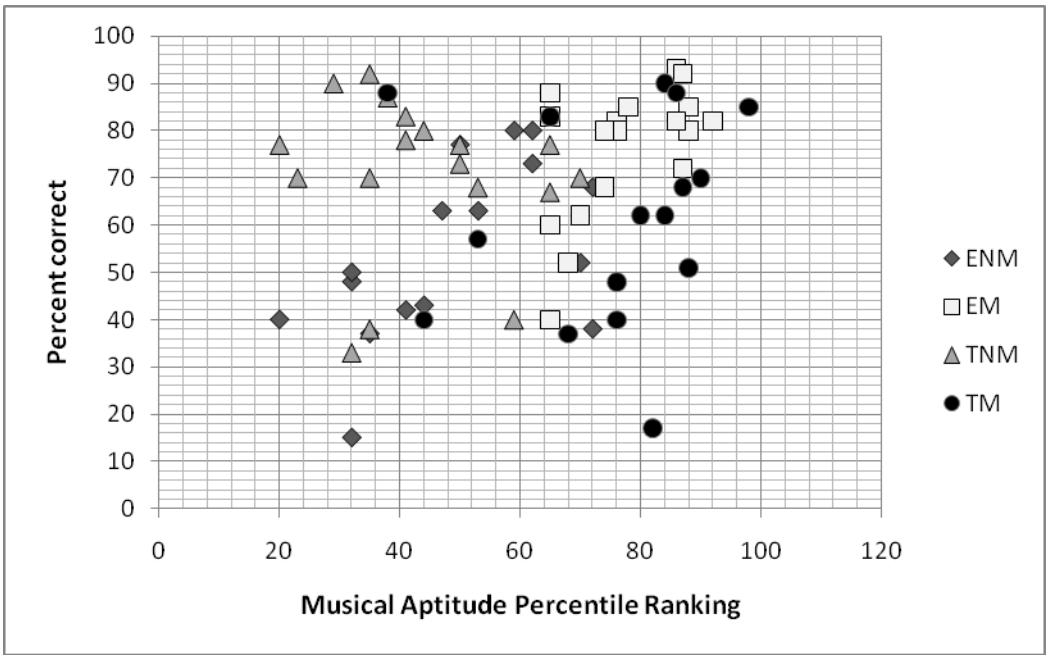


Figure 15 Mean percent correct for Session 7 against the musical aptitude percentile rankings (overall score).

5: DISCUSSION

The discussion section is divided into two main halves. The first half discusses the results of all the main tasks separately, including the pre-/post-training identification task (Section 5.1) and the tone word training program (Section 5.2). Each of these task-based discussions is organized with respect to the main factors involved in the study, that is linguistic and musical experience, and the interplay of these two factors. Section 5.3 is a general discussion that integrates the results from these different tasks to elucidate how the main factors and their interaction impact non-native perception and learning.

5.1 Pre-/post-training identification task

In order to establish participants' level of tone awareness, as well as to determine if tone word training would transfer to improvements in tone identification, a lexical tone identification task was administered before and after training. Results indicated that all groups except the English non-musicians made significant improvements in their lexical tone identification accuracy after tone word training. The improvement made by three of the four groups is consistent with previous studies reporting that perceptual training can lead to significant improvements in identification and discrimination accuracy of non-native tones (Francis et al., 2008; So, 2006a; Wang et al., 1999; Wayland & Li, 2008). It should be noted that these previous studies employed a variety of methodologies specifically focused on lexical tone identification and discrimination training, rather than tone word learning. The results of the present study suggest that even though listeners did not spend the majority of training listening to minimal tone quintuplets, where their

attention would be drawn to the tonal distinctions, nor did they focus on identifying the tones individually, repeated exposure to various tone exemplars aided in their tone category formation. (Davidson, Shaw, & Adams, 2007) suggested that supplying listeners with meaning may provide greater incentive to attend to the relevant phonetic details. One possible explanation as to why English non-musicians, in contrast with English musicians and both Thai groups, did not improve in their tone identification accuracy is that lacking any significant musical or linguistic pitch experience, these listeners likely had a more difficult time tuning into the relevant pitch information, particularly when not presented in minimally contrastive sets. While previous studies have reported significant performance improvements after tone training for naïve English non-musicians (e.g. Francis et al., 2008; Wayland & Li, 2008), tone word training adds another layer of difficulty for the English non-musicians, as there is an additional semantic component to which they need to attend. The English musicians and the Thai groups, on the other hand, were able to utilize their existing pitch experience to aid in the formation of pitch categories over the course of training.

5.1.1 The role of linguistic experience

Despite the fact that Thai listeners made a significant improvement in their scores from pre- to post-test, there was no significant difference overall between English non-musicians and both Thai groups in performance accuracy averaging across pre- and post-tests (e.g. Figure 7). These results do not confirm 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 & Guion, 2004). However, they are perhaps more consistent with Francis et al. (2008), who found

no significant difference in accuracy between Mandarin and English listeners identifying Cantonese tones on either the pre- or post-test. They suggested that group differences in native categories and experience-dependent perceptual cue weightings are reflected in part by differences in tonal confusion patterns. The authors reported that native category representations, particularly for the Mandarin listeners, influenced performance success on certain tones, as these non-native tones assimilated well to the native categories. Furthermore, they also found group differences in tone misidentifications, supporting the notion that weightings of perceptual dimensions in the L1 can affect which dimensions are more salient in non-native tone perception. In the present study, similar results were found in the tone confusion analyses (e.g. Figure 9), which revealed certain group differences that could be attributed to L1 influence. For instance, English non-musicians were misidentifying the Low-Falling tone as Low-Level significantly more than any other tone on both pre- and post-tests; whereas, the Thai non-musicians' misidentifications of the Low-Falling tone were split between Low-Rising, High-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 may stem from their focusing on F0 height. Conversely, Thai listeners confusing Low-Falling as Low-Rising or High-Rising suggests they are 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.

Furthermore, across groups and tests, High-Rising and Low-Rising tones were the most challenging to master. This could be attributed to the fact that both Thai and English possess a rising tone category in each language (lexical and intonational, respectively); thus, the two Cantonese rising tones were likely viewed as different exemplars from a single category, making them more difficult to identify. This is consistent with PAM's "Single Category" assimilation prediction (Best, 1995), which states that two non-native phones produced with similar articulatory gestures as an L1 phone will be difficult to discriminate. These predictions were made primarily for segmental categorization; however, the present results suggest that they can be extended to the suprasegmental domain. Moreover, English listeners were most successful at identifying the High-Level tone, which is logical given that the English intonational system is considered to be comprised of High and Low pitch accent sequences (Beckman & Pierrehumbert, 1986). While one might expect them to be equally successful then for the Low-Level tones, this was not the case. This may have resulted from the fact that the Cantonese low tonal space is more crowded, as it includes Low-Level, Low-Rising, and Low-Falling tones. If English listeners pay greater attention to F0 height rather than contour (Gandour, 1983), this would make it more difficult to discern these subtle contour changes taking place around the same pitch height.

Interestingly, Thai non-musicians were most proficient at identifying Low-Falling and to a lesser extent Low-Level tones across tests. Both tones have a falling contour shape (though Low-Level is a much shallower contour as compared to Low-Falling).

While one might expect that Thai listeners might have a greater challenge if they mapped both low tones to their native Low-Level category, it is conceivable that Thai non-musicians assimilated the Low-Falling to their L1 Falling category and the Cantonese Low-Level tone to their L1 Low category, which may have facilitated their performance on these two tones.

In sum, native group differences were revealed in which tones were easier or more challenging to identify as well as the in their respective tonal confusion patterns, suggesting an influence of native language background in terms of L1 tone categories and perceptual cue weightings on non-native tone identification.

5.1.2 The role of musical experience

The results of the present study confirmed our hypothesis that musical experience would significantly impact lexical tone identification, regardless of L1 background, as both the English and Thai musically-trained participants had higher accuracy rates overall, as compared to their respective non-musician counterparts. Although, English musicians also had significantly higher accuracy rates than the Thai musicians, which may suggest that the L1 tonal background for the Thai musicians was causing some interference. Moreover, given that this was a tone identification task, largely devoid of linguistic information other than tone, musicality may ultimately play a more substantive role than linguistic tonal experience. Schwanhäuβer (2008) similarly reported a musical and language background interaction, in that Australian English musicians were found to be better at discriminating falling tones than Thai musicians. Musical experience became particularly relevant for contrasts such as Low-Rising, Low-Falling and Low-Level tones, which are particularly challenging given their close proximity within the tonal space (e.g.

Figure 1). For the pre- and post-training identification tasks, English musicians were performing significantly better on these tones than non-musicians, suggesting that they were better able to discern the relevant acoustic pitch details. Musicality also appeared to give an edge to the Thai musicians on the pre-test for certain tones, outscoring both non-musician groups on Low-Level identification. By the post-test, Thai musicians also had higher accuracy on High-Rising tones than their non-musician counterparts. Given that pitch is utilized in both music and lexical tone languages to convey meaningful distinctions, the present research points to an overlap in pitch processing mechanisms, as musicians' enhanced pitch sensitivity is beneficial in the linguistic domain. This corroborates previous behavioural research on music and lexical tone identification (e.g. Alexander et al, 2005; Delogu et al., 2009; Gottfried, 2007), which claims that the transferability of musical pitch skills to the linguistic domain counters the notion that music and language are dissociated in processing (e.g. Bever, 1975; Peretz & Coltheart, 2003). These results are also in line with recent neurological research reporting that the processing of non-native linguistic pitch evokes greater mismatched negativity responses, reflecting early, preattentive cortical processing, in English musicians as compared to non-musicians (Chandrasekaran et al., 2009). Similarly, by measuring the frequency following response that encodes the F0 energy in the brainstem, Wong et al. (2007) found more robust encoding in the auditory brainstem of non-native Mandarin tones by English musicians as compared to non-musicians. These neurological findings support the behavioural data of the present study by suggesting that music and language may share some domain-general processing mechanisms.

5.2 Training

5.2.1 Overall Attainment

After seven sessions of perceptual training, listening to Cantonese tone words and learning their associated meanings, all participant groups saw a significant improvement from the first to last session (e.g. Figure 11), increasing their tone word identification accuracy by an average of 42%. This is consistent with previous research training listeners on word meanings that are differentiated by non-native contrasts (e.g. Curtin et al., 1998; Hayes-Harb & Masuda, 2008; Wong & Perrachione, 2007), who reported that non-native listeners were capable of using foreign contrasts lexically, even after a relatively short period of training. These results substantiate the claim that human perceptual systems retain plasticity into adulthood (e.g. Flege, 1995). All groups demonstrated that their perceptual systems were capable of being modified with training, as they learned not only to form new categories for non-native lexical tones but also to use these contrasts to make lexical distinctions.

5.2.1.1 The role of linguistic experience

Our initial hypothesis that Thai non-musicians would attain higher tone word learning proficiency than English non-musicians was confirmed, as Thai non-musicians had greater tone word identification accuracy by the end of training, suggesting an effect of L1 background. The performance asymmetry between Thai and English non-musicians provides general support for the notion that L1 background can have a substantive influence on non-native speech perception (e.g. Best, 1995; Best & Tyler, 2007; Flege, 1995, 2007). Our results suggest that Thai listeners' native language experience with using pitch to differentiate word meaning can beneficially transfer to aid acquisition of

non-native tone words, as Thai non-musicians significantly outperformed English non-musicians on Session 7 tone word identification. 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 that, this may explain why Thai non-musicians, 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, as discussed in Section 2.4.4, 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 (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 aids in their ability to extract the relevant pitch information of non-native words to make meaning associations more efficiently.

5.2.1.2 The role of musical experience

Our hypotheses concerning the influence of musical experience on tone word learning were partially supported. The overall attainment level after training of English musicians was significantly greater than English non-musicians, 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. Moreover, higher musical aptitude percentile ranking scores for both English groups were also found to be significant predictors of word learning success. While musical experience has been shown to facilitate non-native lexical 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 lexical learning, whereby listeners are not asked to identify individual phonemes, but to situate those phonemic contrasts at the word-level to differentiate meaning. These results suggest that musicality (i.e. extensive musical training or aptitude) results in increased auditory acuity, which may aid non-tone language listeners in discerning relevant pitch-cues. This in turn may facilitate the acquisition of the pitch component of these non-native lexical items more efficiently, allowing them to concentrate on mapping the semantic information onto these contrasts. Research has found that musicians possess enhanced verbal memory abilities (Brandler & Rammsayer, 2003; Chan et al., 1998), which may also account for why these musicians were able to attain higher proficiency levels in a tone word learning task.

However, having a musical background did not appear to be particularly advantageous for the Thai musician group, as there was no significant difference between these participants and the Thai non-musicians in overall word learning attainment level. Furthermore, musical aptitude scores for the Thai listeners were not found to be significantly correlated with Session 7 percent correct scores. This is particularly interesting in that it speaks 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 some cases (Slevc & Miyake, 2006), but not in others (Delogu et al., 2009). In the former study, higher musical aptitude in Japanese listeners was correlated with greater proficiency at discriminating and producing challenging English contrasts (e.g. *clown*/*crown*). However, Delogu et al. (2009) reported that Italian musicians did not

significantly outperform non-musicians on a phonological discrimination task in Mandarin Chinese. The authors point out that the majority of Mandarin syllables used in the study contained phonemes common to Italian. This suggests that musicality does not contribute significantly in contexts where listeners are dealing with L2 phonemes that are familiar or even linguistically relevant in their L1. Musicality is perhaps facilitative in cases where listeners are dealing with unfamiliar and difficult L2 contrasts, as in Slevc and Miyake (2006). With respect to the present study, one possible explanation is that musical training may not bear as much influence on tone word learning 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 Thai musicians and non-musicians was found. If the mechanism for pitch to semantic mapping was already established during first language acquisition, then it is conceivable that musicianship would not be able to develop it further. On the other hand, the English musicians did not possess experience with using tone patterns to differentiate word meaning; thus, their musical pitch experience was drawn upon to enhance their ability to utilize linguistic pitch in a higher-level linguistic context.

Furthermore, tonal awareness, either inherent or trained with musical experience, has been shown to impact tone word learning success (Wong & Perrachione, 2007). Thus, to determine whether the ability to identify non-native tones would translate into greater word learning success, a linear regression model was constructed with pre-training identification and session 7 scores as variables. Pre-training tone identification scores were found to be significant predictors of tone word learning proficiency for both

Thai groups and the English non-musicians (e.g. Figure 10). Interestingly, the correlation failed to be significant for English musicians. This lack of correlation for the English musicians may be due to a relatively small standard deviation and high overall mean percent correct for the pre-training identification task scores as compared to the other groups. This limited range in scores in the tone identification task may have made it challenging to account for any variation in the Session 7 results. The fact that tone identification scores predicted word learning success for the other groups is consistent with findings from Wong and Perrachione (2007), who reported that phonetic awareness was a significant factor in word learning. These results suggest that more delineated L2 tone categories will facilitate feature-to-word mappings. It would be much more challenging for listeners to construct lexical representations, with all of the featural components relevant for distinguishing them from other items, and consequently differentiate these lexical items, if the necessary categories to incorporate into these representations are still unstable. Indeed, Weber and Cutler (2004) reported that the confusability of non-native phonemic contrasts leads to the activation of spurious competitors during lexical access, subsequently inhibiting word recognition. They also claimed that more dominant or stable L2 categories, typically those that assimilate better to L1 phonemic categories, have also been attributed to be more dominant for lexical activation.

5.2.1.3 Linguistic versus musical experience

The present research investigated whether linguistic experience facilitated tone word learning to a greater degree than musical experience or vice versa. In order to examine this issue, the comparison between English musician and Thai non-musician

performance was particularly relevant. We originally hypothesized that given the nature of the training task, Thai non-musicians would have an advantage over English musicians, in that they possess greater experience with using pitch to make lexical contrasts. However, the results of the present study did not confirm this hypothesis. In fact, no significant difference in overall attainment level was found between these two groups. Both groups outperformed the English non-musician group, indicating that both of these experiential factors aids tone word learning to some degree, but neither to any greater degree than the other, at least at the initial stage of learning. These findings provide support for the notion that language and music share some cognitive mechanisms (Koelsch et al., 2004; Patel, 2008, 2003). Previous research has pointed to a convergence in music and language processing for other domains of linguistics, namely syntactic processing (Patel, 2003). They suggest that music and language may be dissociated at the level of representation, but overlap for cognitive processing. Similarly, the present study found that possessing either a tone language background or musical experience with a non-tone language background results in significantly better word learning attainment than non-tone language non-musicians. This suggests that enhanced processing mechanisms for pitch developed as a product of sustained linguistic pitch experience (tone language listeners) or long-term musical pitch experience (musicians) may not be domain-specific and can both aid the acquisition of tone words for the initial learning stage, given that Thai non-musicians and English musicians both performed significantly better than English non-musicians. The interaction of these two factors will be discussed in further detail in Section 5.3.

5.2.2 Training Improvement Trajectories

In addition to overall attainment level, the pattern of learning for each group over the course of training was also analyzed in order to investigate any group differences in the speed and degree of improvement. All groups significantly improved from Day 1 to 2 and from Day 2 to 3; although, no significant improvement was made for any group from Day 3 to 4. This may be attributed to the fact that all other days involved two training sessions; whereas, Day 4 only contained one. A significant increase in performance may have been obtained if participants had the opportunity to complete an additional training session on the last day.

There were interesting group differences with respect to their training improvement trajectories (e.g. Figure 13). English musicians performed significantly better than English non-musicians and Thai musicians by the second day of training. However, it was only on the final training day that Thai non-musicians were significantly better than English non-musicians, and the Thai musicians were no longer significantly worse than English musicians or Thai non-musicians. It is possible that the Thai groups' more gradual pace of improvement as compared to the English musicians results from interference of their L1 prosodic system. This slower improvement rate may have been a reflection of the fact that these non-native lexical tones needed first to be processed with respect to an existing lexical tone inventory, similar to the processing of non-native segmental information (e.g. Best, Halle, Bohn & Faber, 2003). Flege's Speech Learning Model (1995) posits that L2 phonemes that are new or phonetically dissimilar from L1 phonemes should be easier to acquire than those that are similar to existing native phonemes. In the present study, Thai listeners may have attempted to map L2 tones to

their existing native tone categories, which may have contributed to their more gradual progress. For example, the presence of several L2 tones, such as High-Level, High-Rising, Low-Level and Low-Falling, which may share perceived phonetic similarities to Thai tones (e.g. High, Rising, Mid, Low), may have made it more challenging initially for Thai listeners to form L2 categories. While one may wonder why this L1 prosodic interference did not slow the English musicians' progress, as the English prosodic system possesses intonational categories, it has been suggested that the lower functional load of intonation, as compared to lexical tone, may cause them to exert a weaker influence on incoming non-native tones (Francis et al., 2008). English intonation is considered to have a lower functional load than lexical tone because it is less linguistically meaningful, in that it is not used phonemically. This weaker influence may have allowed English musicians to more efficiently form new tone categories for these non-native tones, thereby speeding up their improvement rate, as compared to the Thai groups.

5.2.3 Error type

Error type analysis was performed in order to determine if the errors made resulted from misidentifying the tonal or segmental components of a lexical item. This analysis found significantly more tonal errors than segmental errors made at both the beginning and end of the training program for all groups. This is consistent with Wong and Perrachione (2007), who found that listeners were making primarily tonal errors by the end of training. Furthermore, the proportion of tonal errors significantly increased from the first to last session, with concurrent decreases in segmental and segmental + tonal (both) errors across groups. Given their proficiency with pitch, one might wonder why musicians would also be making primarily tonal errors, even by the end of training.

However, previous research suggests that listeners, even native ones, process segmental information much faster and with more accuracy than tonal information (Cutler & Chen, 1997). These results highlight the fact that by the end of training, despite significant improvements for most groups in tone identification, the tonal component was the most challenging aspect for all listeners to acquire. Those who made significantly more errors (i.e. English non-musicians) were thus making almost exclusively tonal errors, suggesting that acquisition of the tonal component of the lexical items is attained at a later stage in learning, after the segmental component. This notion is consistent with results reported in Curtin et al. (1998), who suggested that non-native listeners first create lexical representations derived from the features that are lexically contrastive in their L1. In the case of the present study, segmental information appeared to be acquired first, which is logical given that segmental contrasts are lexical in both English and Thai. Furthermore, there were only three syllables used in training, which also would have been comparatively easier to remember as compared to five non-native tones.

Error type analysis also provided a window into any group differences in how lexical representations are constructed. Interestingly, English musicians had a significantly higher proportion of segmental errors than all other groups overall. This finding diverges from Wong and Perrachione (2007) who reported no significant difference in the proportions of error types between the successful (largely comprised of musicians) and less-successful learner groups. One possible explanation is that the musicians in this study focused their attention on learning the tonal information across training sessions, as this component may have been more perceptually salient (particularly to musically-trained listeners). The other groups may have conversely honed

in on segmental information to a greater degree, thereby reducing their segmental errors in comparison to the musicians. Pitt (1994), using musical stimuli, reported that non-musicians tended to weight the timbre (spectral) dimension more heavily than pitch and found it more challenging to discern pitch differences when timbre also varied. In other words, timbre was more perceptually salient for non-musicians. However, musicians were found to be able to process each of these dimensions (pitch and timbre) independently.

Additionally, Thai non-musicians were found to have significantly more tonal errors than English and Thai musicians. This is not surprising given that the musician groups demonstrated higher performance in the tone identification task. Musical experience may have enabled musician listeners to make fewer tonal errors overall. It should be noted that these results averaged across the first and last sessions. The majority of this difference between Thai non-musicians and the musician groups was likely derived from the first session (though not significantly), as the difference in tonal errors between the English musicians and Thai non-musicians was 17%, as compared to the last session where this difference was reduced to 3%. The reduction of tonal errors by the last session for Thai non-musicians is likely attributed to their establishing more stable tonal categories from repeated exposure over the course of seven sessions, as reflected by their significant improvement from pre- to post-training tone identification.

5.3 General Discussion

The primary goal of the present study was to elucidate the relative and combined influences of linguistic and musical experience on Cantonese tone word learning. Previous research has primarily focused on how these factors impact non-native lexical

tone identification separately (e.g. Alexander et. al, 2005; Francis et. al, 2008; Gottfried, 2007; Wayland & Guion, 2004); however, the role of both these factors in how listeners utilize non-native lexical tone in a more linguistic context has not been fully investigated. The interplay of these factors on tone word learning has interesting ramifications on the nature of the mechanisms involved in category identification and feature-to-word mapping. This discussion section will address how the results from the different tasks employed in this study come together to provide a window into the complex role of linguistic and musical experience, as well as their interaction, on non-native perception and learning.

5.3.1 Linguistic experience

Native language background can potentially facilitate or inhibit second language perception (e.g. Aoyama et al., 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 PAM (e.g. Best, 1995; Best & Tyler, 2007) and SLM (e.g. Flege, 1995, 2007). They posit that some non-native contrasts will prove more challenging than others, depending on the degree of perceived perceptual or gestural similarity between the L2 contrast and the L1 category. While these theories have primarily focused on segmental distinctions, some studies have provided evidence that native tonal categories may also influence the perception of non-native tones (e.g. Francis et al., 2008; So, 2006b; Wang, 2006). The results of the current study illustrate that the interaction of L1 and L2 may have differing effects, depending on a variety of factors including task type and stage of learning. Indeed, the present research found that tone language experience may be facilitative in some contexts, such as at the word-learning level, but not necessarily at the level of

phonemic identification, as demonstrated by the pre-/post-training tone identification scores.

With regards to non-native tone perception, it appears that tone language experience is not necessarily more advantageous for identifying non-native lexical tones, as both English and Thai non-musicians had similar performance accuracy across pre- and post-training identification tests. While these results are contra studies such as Burnham et al, (1996), Wayland and Guion (2004) and Lee et al. (1996), they are in line with previous research reporting that a tone language background does not necessarily facilitate (Francis et al., 2008) and can even inhibit non-native tone perception (So, 2006b; Wang, 2006). Consistent with findings from Francis et al. (2008), the results of the pre- and post-training identification task in the present study suggest an influence of L1 background and tone categories, which manifested in differing tonal confusion patterns and which tones were more challenging for listeners. For instance, English listeners performed best on High-Level tone identification; whereas, Thai listeners had better performance on Low-Falling and Low-Level tones. There was also evidence to suggest an L1 influence on the weightings of certain perceptual dimensions, such as F0 height and direction of change (Gandour, 1983). The relative weights of these perceptual features for each listener are established as a consequence of tone categorization in their respective L1s. English listeners are purported to attend to F0 height more than direction; whereas, tone language listeners have been found to focus on direction. The results of the present research provide some evidence in support of these claims. However, it should be noted that suggestions regarding L1/L2 category mapping remain speculative, in the absence of a direct acoustic comparison of the prosodic systems or a similarity/goodness

of fit judgment task. These observations merely serve to illustrate that native language background appears to have a significant influence on non-native tone perception.

Furthermore, the interaction of the L1 and L2 tonal systems also appeared to influence the improvement progress of tone word learning to a certain degree, particularly at onset of learning. Indeed, the training improvement trajectory for each group illustrated that it was not until the end of training that the Thai groups were significantly better than the English non-musicians. One explanation for this is that Thai listeners may possess more robust native tone categories with which the incoming non-native tones need to negotiate (Halle, Chang, & Best, 2004). As discussed in Section 5.2.2, the construction of new tone categories in relation to existing ones, such as the Cantonese High-Level, High-Rising and Low-Rising tones negotiating with Thai High and Rising tones, may have consequently slowed their improvement progress. This was reflected in their lack of significant improvement in tone word identification accuracy over the first three days, relative to the English musicians, who saw significantly greater improvement early in training.

While tone language experience may have been somewhat inhibitory at the earliest word learning stage when learners are initially forming appropriate non-native tone categories, slowing the rate of improvement, the results of this study indicate that it was facilitative when it came to overall attainment level in tone word learning. In particular, the experience of using native tone distinctions to make lexical contrasts significantly aided the ability to use non-native tones to differentiate word meaning, as Thai non-musicians reached a significantly higher level of attainment than the English non-musicians. These behavioural results are in line with neurobiological data suggesting

that mechanisms in the brainstem for representing pitch, as manifested by pitch-tracking accuracy, are more receptive in tone language speakers than non-tone language speakers, and that this pitch encoding ability is transferable to non-native tone languages (Krishnan et al., 2010). The similarity in pitch-tracking accuracy patterns in the auditory brainstem between Mandarin and Thai groups, regardless of whether the tone bore linguistic relevance in the listeners' L1, suggests that tone language listeners have enhanced lower-level pitch sensitivity. One may wonder why such superior pitch encoding in the auditory brainstem, as demonstrated in Krishnan et al. (2010), did not also equal greater pre- or post-training tone identification proficiency for the Thai listeners. One possible explanation may involve the nature of the tone identification and tone word learning tasks, as the former involves assigning specific labels to tone categories, while the latter involves hearing tone words and associating meanings to them. The latter may be more akin to how Thai listeners acquire new lexical items, processing both segmental and tonal information and mapping that unit onto a concept, rather than singling out and naming an individual component (tone) of that item. This enhanced neurobiological pitch encoding capacity demonstrated in tone language speakers in previous research (Krishnan et al., 2010) may be reflected in their ability to use tone to make lexical distinctions rather than aid in their ability to assign tone labels.

The findings for the tone word learning task could be interpreted with respect to the native language functional load of pitch (Gandour et al., 2002; van Lancker, 1980). Functional load can be defined in terms of the frequency of occurrence as well as the level of contrastivity (King, 1967), that is, how many minimal pairs with a given contrast exist in the language, as well as to what degree these pairs are contrastive (i.e. lexical,

grammatical, emotional). English utilizes pitch with relatively low functional load, in that stress and intonation are used to mark grammatical contrasts or denote pragmatic or emotive information. On the other hand, pitch in Cantonese and Thai has high functional load and is the "most systematically linguistic" (van Lancker, 1980), as it is used to make lexical contrasts on all words. It is perhaps more challenging for listeners to acquire words where there is an L1-L2 disparity in functional load for a contrast, particularly when they are required to shift from low to high functional load (McAllister et al., 2002). These listeners need to learn not only to attune to cues that hold less linguistic significance in their native language but also to apply them to make lexical contrasts.

In sum, it appears that different aspects of language background are drawn upon during different stages of acquisition. L1 categories and native weightings of perceptual dimensions influence the categorization of non-native phones, which, depending on the specific nature of the phonetic L1 and L2 inventories, may not be facilitative (e.g. Francis et al., 2008). In this study, Thai and English non-musicians performed comparably on the pre-/post-training tone identification task, regardless of the function of pitch cues in their native languages. However, in the word learning domain, experience with its function as being lexically contrastive in Thai became advantageous for Thai non-musicians over their English counterparts, perhaps because Thais had more experience with incorporating pitch information into lexical representations.

5.3.2 Musical experience

The present study found that the role of musicality differed depending on language background (tone versus non-tone language) and the nature of the task (phoneme tone identification, tone word learning). English musicians, who possessed

more musical training as well as obtained higher musical aptitude scores than non-musicians, were found to perform significantly better on both pre- and post-training identification tasks as well as on the tone word learning tasks, consistent with results of previous studies (e.g. Alexander et. al, 2005; Wong & Perrachione, 2007). Musical training was similarly helpful for Thai listeners in the tone identification task, in that they achieved higher identification accuracy than the Thai non-musicians. However, it did not significantly enhance their overall attainment level on the tone word learning task as compared to their Thai non-musician counterparts. One possible explanation for this discrepancy between Thai musicians and non-musicians on the pre-/post-training identification task is that musicians are perhaps more accustomed to attaching labels to tone categories, given the naming of pitch intervals in music (Patel, 2008), which may have aided in their assignment of tone category labels in the identification task.

One might assume that the Thai musicians would have achieved a higher overall attainment level on the tone word learning task than the Thai non-musicians, given that they performed significantly better at identifying non-native lexical tones, and that pre-training tone identification scores significantly predicted word learning success. However, it should be noted that tone identification only accounted for about 34% of the variance in attainment level in our regression model, indicating that other factors, such as feature to word mapping abilities, may also bear a significant role in word learning success. Another explanation is that word learning tasks, for both first and second language development, have been found to increase processing demands that may make it challenging to access newly-acquired, unstable contrasts (Hayes-Harb & Masuda, 2008; Pater, Stager, & Werker, 2004; Stager & Werker, 1997). Hayes-Harb and Masuda (2008)

found that English learners of Japanese did not perform as well as might be expected on a lexical test (i.e. auditory word-picture matching task), given that they were able to discriminate the non-native contrasts (singleton vs. geminate consonants) with a high degree of accuracy (*M*=93%). They attributed the diminished accuracy rate on the lexical test to increased processing costs incurred by the word learning task. This may explain why Thai musicians were able to perform better than non-musicians on a pre-training identification task, but to not have that success translate into greater proficiency on a more cognitively demanding task.

The fact that Thai musicians did not achieve a significantly higher overall attainment level for tone word learning than the Thai non-musicians suggests that the combination of musicianship and tone language experience may not be additive. Musical experience facilitated tone word learning for listeners without a tone language background (English musicians versus non-musicians), but no such facilitating effect was found for those with a tone language background. This points to an influence of L1 background, whereby musical pitch experience does not appear to further develop existing tone to word association mechanisms or further enhance the weight of the linguistic pitch dimension, whose salience may have already been increased as a result of tone language experience.

5.3.3 Linguistic versus musical experience

Despite English musicians' demonstrated pitch acuity, Thai groups were hypothesized to have greater success in tone word learning because of their experience with using pitch contrasts systematically to convey lexical distinctions. However, no significant differences were found between the English musician and both Thai groups by

the last session of tone word training. Curtin et al. (1998) suggested that listeners first construct lexical representations based on features that are lexically contrastive in their L1. As discussed in Section 5.3.1, this prediction held for Thai versus English non-musicians, as experience with employing pitch to make lexical distinctions resulted in significantly better performance for the Thai non-musicians. Thus, one might ask why English musicians would achieve comparable proficiency as the Thais in word learning, considering that tone is not used to make lexical distinctions in English. One explanation is that their musical expertise and experience with pitch may facilitate the elevation of the linguistic pitch dimension to such a degree that they are able to identify the non-native tones as being behaviourally relevant and incorporate that information into their lexical representations, despite it not being contrastive in English.

Recent research may provide a neurobiological explanation for the similar overall attainment levels achieved by the English musician and Thai groups, 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., 2009) and musicians (Wong et al., 2007). Both studies posit that the sensitivity of brainstem neurons that extract relevant pitch information is enhanced by experience-dependent subcortical mechanisms. The current findings provide behavioural evidence supporting the idea that this domain-general subcortical tuning, resulting from either musical or tone language experience, can in fact aid tone word learning, in that it may facilitate a more efficient process of associating pitch information with semantic content.

Some researchers have proposed a modular approach to music processing, as distinct from linguistic, suggesting that processing resources are shared at an early stage

of acoustic analysis before diverging into distinct language and music modules of processing (Peretz & Coltheart, 2003). However, Wong and Perrachione (2007) suggested that this divergence may in fact occur after the acoustic analysis stage, and that this overlap may include sound association and memory formation. The tone word learning results of the present research provide evidence of this larger domain of shared processing. To learn the meanings of the lexical items in this study, participants needed to learn both components of the word (segmental and tonal) and map them to the semantic information appropriately. While tone language listeners have accumulated experience with this process, non-tone language listeners do not typically break down these components and associate both aspects with meaning. For instance, Cutler (1986) noted that English listeners typically normalize across prosodic information for the purposes of lexical access, as segmental information is considered sufficient information to retrieve the appropriate lexical item. Additionally, word recognition is impaired to a greater degree when mis-stressing results in vowel quality changes (e.g. *wallET, DEceit*) over simply inappropriate pitch placement (Cutler & Clifton, 1984). However, the fact that the English musicians performed equally well as the Thai listeners on the tone word learning task, both significantly outperforming the English non-musicians, suggests that this mechanism of mapping concepts to pitch categories may be more domain-general, at least for the initial stage of learning. Both musical and tone language experience appear to develop the perceptual weight of the pitch dimension, which may subsequently aid in more cognitively demanding tasks, such as learning sound-concept associations.

5.3.4 Summary: The complex role of linguistic and musical experience on second language learning

One of the major questions of this research was to examine whether linguistic and non-linguistic (musical) factors would facilitate word learning in Cantonese. The findings of the present study suggest that the influence of these factors appears to be in part dependent on linguistic context (e.g. phonetic versus word level) as well as L1 experience with the functional use of pitch. Functionality appears to interact with linguistic context, in that the L1 status of pitch did not have a significant influence for the tone identification task, but played more of a role at a higher linguistic context (word learning). While there is a discrepancy between English and Thai with respect to the degree of the functional load of pitch, English and Thai non-musicians did not differ significantly at identifying non-native Cantonese tones overall. Indeed, it did not appear to bear on the identification performance of the English and Thai musicians, with English musicians performing significantly better on tone identification than both Thai groups. However, experience with its function as lexically contrastive was advantageous at the tone word learning level, with Thai non-musicians significantly outperforming English non-musicians by the end of training.

The role of musicianship was also influenced by task, context and linguistic background. Long-term musical pitch experience was particularly advantageous on a tone identification task. Such a task potentially abstracts away from linguistic information and focuses largely on F0 modulations (which for many musicians could be viewed as a musical task, matching pitch contours to the appropriate pitch diagrams). With respect to word learning, as illustrated in the discussion in Section 5.3.2, the combination of tone language and musicality appears not to be additive, in that the combination did not

produce significantly better results in tone word learning than a tone language alone. This may also return to the notion of functionality, in that pitch distinctions have a lexical function in the L1 for both the Thai musicians and non-musicians. Consequently, musicality may not be able to provide additional aid for the tone language listeners. However, musical experience appears to enhance pitch to word association mechanisms that have not been fully developed by tone language experience, such as for the English musicians, aiding listeners in overcoming this functional asymmetry.

Taken together, the results of the present research illustrate that the influence of linguistic and musical experience may be modulated by linguistic context, 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 levels and stages. At a lower-level context, such as tone identification, L1 phonetic inventories and perceptual cue weightings play a role, manifesting in differing tonal confusion patterns for example. Yet, it was musical experience that significantly influenced overall success on tone identification, rather than prior experience with discerning linguistic pitch distinctions. However, there are different cognitive requirements for listeners at a higher-level linguistic context (i.e. word learning), which subsequently shifts what these factors contribute. Prior experience with *utilizing* linguistic pitch in a lexically significant manner was useful at later stage in the learning process, allowing Thai non-musicians to achieve word learning success faster than English non-musicians. Musicality appears to be a significant factor at this level only when L1 experience has not previously developed the appropriate mechanisms involved in feature to word mapping. These findings suggest that linguistic and non-linguistic factors have dynamic roles in the L2 learning process,

and that their influences may vary depending on the needs of the context. Moreover, the present research points to an overlap between music and language processing mechanisms, at both the tone and tone word level. The implications for a theory of L2 learning and perception could be that mechanisms involved in L2 learning, at least at the initial stages, may not be specific to the linguistic domain and are susceptible to influence from other experiential factors.

6: CONCLUSIONS & FUTURE DIRECTIONS

In sum, the findings of the present research can be considered with respect to the three main streams outlined in Section 2.6.2, namely linguistic experience, musical background and the interaction of these factors. First, tone (Thai) and non-tone (English) non-musicians were found to not differ appreciably in performance on the pre-/post-training tone identification task, suggesting that tone language experience was not necessarily advantageous for non-native tone identification. However, the nature of their respective L1 tonal inventories and perceptual cue weightings provided some insight into group performance discrepancies on specific tonal contrasts, consistent with previous speech learning theories and empirical findings (e.g. Flege, 1995, 2007; Francis et al., 2008; Gandour, 1983).

Linguistic background and its interaction with musical experience did appear to affect tone word learning proficiency. Native experience with the lexically contrastive function of pitch provided an advantage for the Thai non-musicians, who outperformed English non-musicians at tone word identification by the end of training. Furthermore, Thai listeners with musical experience did not display significant performance differences from their non-musician counterparts in overall attainment level, suggesting that these factors are not additive at the initial stage of learning, and that existing tone to word association mechanisms developed during L1 acquisition are not further enhanced by musical training. Finally, musical experience benefited English listeners, enhancing the weight of the pitch dimension, both musical and linguistic, and facilitating tone to

word mapping. These findings point to an overlap between music and language processing, in that the cognitive mechanisms utilized to associate pitch to semantic information appear to be domain-general, differentially developed as a product of experience.

While the findings of the present study have provided some insight into the complex roles of musical and linguistic experience on non-native perception, there are still several avenues of future research to pursue. First, it should be stated that linguistic and musical experience are by no means the only factors that impact tone word learning. The experimental design of this study allowed for a controlled examination of these specific factors. The participant groups were largely comparable with respect to age and educational background, and randomly-sampled in order to potentially balance any extraneous factors. However, a separate investigation into factors such as language aptitude and motivation would also be interesting, particularly considering most previous research has dealt with non-tonal languages. While Schwanhäuβer (2008) reported that foreign language aptitude alone was not a predictor of tonal or segmental perception, it is conceivable that it could play a role at a higher linguistic level, such as word learning. Indeed, attitude, motivation and language aptitude have been found to influence how quickly second language material is acquired (Gardner, Lalonde, & Moorcroft, 1985). Additionally, it may also be interesting to examine how factors such as cognitive learning styles can account for individual learner variation in tone language learning (e.g. Ehrman & Leaver, 2003).

Second, Wong and Perrachione (2007) noted that employing a performance-based training termination criterion allows learners to reach their highest attainment levels and

provided a clearer picture of the overall learning process. The present study utilized a time-based criterion, with a fixed set of sessions. Some less-successful learners in the Wong and Perrachione (2007) study took upwards of 18 sessions to reach criterion, and the potential for prolonged training programs was not feasible for the present study. However, these time-based training programs could be construed as more akin to language learning in a classroom setting, in that learners are often required to learn a set of vocabulary items within a fixed period of time. The present study also controls for the amount of input learners received. However, future studies could follow learners through a longer duration of training to examine how linguistic and musical experience influence the tone word learning process through later stages of acquisition.

It would also be useful to investigate how these factors play a role in broader linguistic contexts. The present study investigated the initial period of learning within a single-word context. It is conceivable that as the L2 develops, incorporating other domains of linguistic information such as prosodic and syntactic structure, language background (e.g. tone versus non-tone) may play a more influential role than musical experience. 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).

Finally, for non-tone language non-musicians, it would be interesting to determine if focused lexical tone training prior to undertaking the tone word learning program would produce similar tone word proficiency results produced by the English musicians in the present study. Tone training could include repeated exposure to minimal tone pairs or practice with identifying and discriminating tonal contrasts (e.g. Wang et al., 1999).

This may point to whether it is in fact just enhanced pitch acuity, or if there are other aspects, such as overall intelligence or fine motor skills, developed by musical experience that may help with word learning (Schellenberg, 2003). If tone training prior to the tone word learning program enables non-musicians to reach similar attainment levels as the musician group in tone word learning, this would suggest that improved tonal perception is the primary aid in the word learning process. However, if the non-musicians fail to reach similar attainment levels, this may imply that some other aspect of musical training contributes to word learning proficiency. The fact that the musical aptitude scores were positively correlated with tone word learning attainment level for the English listeners in the present study suggests that such lexical tone training may provide a significant boost to non-musician word learning scores, as musical aptitude measures auditory abilities and does not involve any kind of other experiential factors. Indeed, Song, Skoe, Wong, and Kraus (2008) found that after undergoing a lexical identification training program, similar to Wong and Perrachione (2007), English non-musicians demonstrated improved pitch-tracking accuracy in the auditory brainstem. Thus, focused lexical tone training may be able to improve such accuracy to an even greater degree.

Thus, the results of the current study have provided additional evidence pointing to an overlap in the cognitive mechanisms involved in processing music and language, not only at the level of phonetic discrimination but also at a higher linguistic level such as word learning. However, this research also highlights the fact that the influences of linguistic and musical experience are by no means straightforward, with issues such as functional load and learning stage (e.g. phoneme categorization, word learning) interacting with these factors and affecting their relative contributions during learning.

APPENDICES

Appendix A: Participants' musical backgrounds

English group: Non-musicians (those without any experience have been omitted)

	INSTRUMENTAL EXPERIENCE			VOCAL EXPERIENCE	
Subject	**Instrument(s)**	**Age started**	**Duration (yrs)**	**Age Started**	**Duration (yrs)**
102	recorder	8	4	-	-
104	guitar/bass	9/12	0.1/0.8	-	-
105	clarinet	13	1	-	-
106	saxophone/guitar	11/20	3/0.5	10	1
108	recorder/alto saxophone	10/10	1/2.5	10	1
112	guitar	17	1	-	-
115	clarinet/piano	12/7	4/2	-	-
118	piano	9	2	-	-

English group: Musicians

	INSTRUMENTAL EXPERIENCE			VOCAL EXPERIENCE	
Subject	**Instrument(s)**	**Age started**	**Duration (yrs)**	**Age Started**	**Duration (yrs)**
201	guitar/piano	16/21	2/1	5	18
202	piano/trumpet	5/11	3/16	7	20
203	piano/drum/guitar	7/21/17	7/1/1	6	16
204	piano.trumpet/euphonium/musical saw/guitar	7/11/14/17/17	19/3/12/9/9	17	5
205	piano/flute	5/13	15/7	18	2
206	piano/guitar	5/17	19/5	5	19
207	piano/trombone/guitar/euphonium	7/14/15/13	13/7/6/8	17	3
208	piano/flute	3/12	7/3	3	22

209	oboe/flute/piano	14/12/8	4/2/10	8	10
210	violin/piano/guitar	7/12/16	12/7/3	10	4
211	piano/oboe/auxiliary percussion	4/11/14	17/4/4	16	5
212	guitar/drum kit/ percussion	13/17/18	9/5/4	16	6
213	piano/saxophone	6/12	9/1	7	18
214	clarinet/saxophone	12/13	14/13	8	6
215	piano/guitar	5/15	10/4	9	10
216	piano/trumpet/French horn/guitar	7/12/13/15	11/6/5/3	12	4
217	piano/guitar/clarinet	9/25/12	21/2/5	7	22
218	guitar/piano	16/18	10/8	12	14

Thai group: Non-musicians (those without any experience have been omitted)

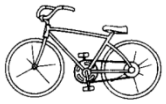
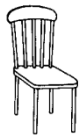
	INSTRUMENTAL EXPERIENCE			**VOCAL EXPERIENCE**	
Subject	**Instrument(s)**	**Age started**	**Duration (yrs)**	**Age Started**	**Duration (yrs)**
302	Kim (Thai string instr.)	10	2	15	1
303	Flute	17	1	17	1
305	guitar	10	3	-	-

Thai group: Musicians

	INSTRUMENTAL EXPERIENCE			**VOCAL EXPERIENCE**	
Subject	**Instrument(s)**	**Age started**	**Duration (yrs)**	**Age Started**	**Duration (yrs)**
403	violin	13	7	-	-
405	trombone	13	6	-	-
409	cello	13	9	-	-
410	viola/violin	17/12	3/6	-	-
411	cello/violin	13/20	12/5	-	-
412	violin	12	7	-	-
413	piano/drums/percussion	6/15/10	10/8/10	-	-
414	piano	4	17	21	1
415	violin	11	7	11	7
416	piano	13	7	-	-

417	violin/viola/keyboard/sax	13/15/13/16	7/5/7/4	-	-
418	piano/drums/trumpet	12/16/20	7/2/6	12	8
419	guitar/bass	10/23	14/1	10	8
420	clarinet/sax	12/16	7/3	-	-
421	bassoon	13	9	-	-
422	percussion	13	9	-	-

Appendix B: Training stimuli

	/tsou/	**/k^{w}aj/**	**/wu/**
Tone	**Assigned meaning**		
High level (55)	 “umbrella”	 “fish”	 “bicycle”
High rising (25)	 “hand”	 “book”	 “chair”
Low falling (21)	 “frog”	 “leg”	 “scissors”
Low rising (23)	 “watch”	 “cat”	 “ear”
Low level (22)	 “key”	 “rat”	 “glasses”

Appendix C: Pre-/post-training identification stimuli

Chinese character	Syllable	Tone
威	/waj/	High level (55)
位		High rising (25)
維		Low falling (21)
偉		Low rising (23)
慧		Low level (22)
嚕	/low/	High level (55)
佬		High rising (25)
牢		Low falling (21)
老		Low rising (23)
路		Low level (22)
詩	/si/	High level (55)
史		High rising (25)
時		Low falling (21)
市		Low rising (23)
是		Low level (22)
丕	/pej/	High level (55)
鄙		High rising (25)
疲		Low falling (21)
被		Low rising (23)
贔		Low level (22)
呼	/fu/	High level (55)
虎		High rising (25)
扶		Low falling (21)
婦		Low rising (23)
負		Low level (22)

Appendix D: Confusion matrices

TNM Pre-test

Token	Identified as				
	55	25	21	23	22
55	31.7	13.6	5.8	16.4	21.4
25	31.7	31.4	6.7	15.8	6.1
21	5.0	12.2	48.6	17.5	13.1
23	19.4	28.3	16.1	28.1	16.4
22	8.3	10.6	20.8	19.2	40.3
No Resp	3.9	3.9	1.9	3.1	2.8

TM Pre-test

Token	Identified as				
	55	25	21	23	22
55	53.0	2.7	2.0	8.7	9.3
25	22.3	44.0	3.0	15.7	3.0
21	2.7	5.7	55.3	10.0	12.7
23	11.0	40.3	21.7	40.3	9.0
22	10.0	4.7	15.3	23.0	63.0
No Resp	1.0	2.7	2.7	2.3	3.0

ENM Pre-test

Token	Identified as				
	55	25	21	23	22
55	66.9	4.4	3.4	6.9	20.0
25	12.8	39.1	7.8	15.3	5.0
21	5.9	7.5	45.6	9.7	12.8
23	6.9	34.4	8.8	39.7	14.7
22	5.9	12.2	31.9	24.7	43.1
No Resp	1.6	2.5	2.5	3.8	4.4

EM Pre-test

Identified as	Token				
	55	25	21	23	22
55	94.2	0.0	0.6	2.5	13.6
25	0.0	66.1	0.3	16.4	0.0
21	0.0	0.6	67.2	2.5	7.2
23	0.8	32.5	2.5	66.4	4.2
22	4.4	0.3	29.4	11.4	73.9
No Resp	0.6	0.6	0.0	0.8	1.1

TNM Post-test

Identified as	Token				
	55	25	21	23	22
55	53.6	8.3	2.8	14.4	12.5
25	26.4	39.4	5.8	16.9	4.2
21	1.4	4.7	67.5	7.5	6.4
23	12.8	33.9	10.3	43.1	11.1
22	3.6	6.7	10.8	12.8	61.1
No Resp	2.2	6.9	2.8	5.3	4.7

TM Post-test

Identified as	Token				
	55	25	21	23	22
55	72.7	4.3	0.7	6.0	10.7
25	12.0	58.3	3.0	17.7	2.7
21	1.0	3.7	75.7	6.0	7.3
23	7.3	24.3	7.0	56.3	10.3
22	5.0	5.7	9.3	8.7	62.7
No Resp	2.0	3.7	4.3	5.3	6.3

ENM Post-test

Identified as	Token				
	55	25	21	23	22
55	73.8	4.4	0.9	3.8	16.9
25	10.0	47.2	2.5	16.6	3.8
21	0.9	3.1	56.3	5.9	7.2
23	6.9	32.2	5.9	44.4	15.3
22	7.8	9.4	30.6	25.3	52.8
No Resp	0.6	3.8	3.8	4.1	4.1

EM Post-test

Identified as	Token				
	55	25	21	23	22
55	96.4	0.0	0.0	0.0	11.7
25	0.0	78.9	0.6	18.1	0.0
21	0.6	0.0	85.3	0.6	2.8
23	0.0	20.0	0.3	75.3	0.6
22	2.8	0.3	13.6	4.7	84.2
No Resp	0.3	0.8	0.3	1.4	0.8

REFERENCE LIST

Abramson, A. S. (1962). The vowels and tones of standard Thai: Acoustical measurements and experiments. *International Journal of American Linguistics*, *28*(2).

Abramson, A. S. (1975). The tones of central Thai: Some perceptual experiments. In J. Harris & J. R. Chamberlain (Eds.), *Studies in Tai Linguistics. In honor of William J. Geddney*. Bangkok: Central Institute of English Language.

Abramson, A. S. (1976). Thai Tones as a Reference System. In T. Gething, J. Harris, & P. Kullavanijaya (Eds.), *Tai Linguistics in Honor of Fang-Kuei Li*. Bangkok: Chulalongkorn University Press.

Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, *21*(4), 319-325.

Abramson, A. S. (1979). The coarticulation of tones: An acoustic study of Thai. In T. Thongkum, P. Kullavanijaya, V. Panupong, & K. Tingsabadh (Eds.), *Studies in Tai and Mon-Khmer phonetics and phonology in honour of Eugenie J. A. Henderson*. Bangkok: Chulalongkorn University Press.

Alexander, J. A., Wong, P. C., & Bradlow, A. R. (2005). Lexical Tone Perception in Musicians and Non-musicians. In *Proceedings of Interspeech 2005*. Presented at the Eurospeech – 9th European Conference on Speech Communication and Technology., Lisbon, Portugal.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004).

Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, *32*(2), 233-250. doi:10.1016/S0095-4470(03)00036-6

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese Phonology*. Trends in linguistics. Berlin: Mouton de Gruyter.

Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 255-309.

Besson, M., & Schön, D. (2001). Comparison between language and music. *Annals of the New York Academy of Sciences*, *930*, 232-258.

Best, C. T. (1994). Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Mode. In J. C. Goodman & H. C. Nusbaum (Eds.), *The Development of Speech Perception the Transition from Speech Sounds to Spoken Words*, MIT cognet (pp. 167-224). Cambridge, Mass: MIT Press.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Baltimore: York Press.

Best, C. T., Halle, P. A., Bohn, O., & Faber, A. (2003). Cross-language perception of nonnative vowels: Phonological and phonetic effects of listeners' native languages. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2889-2892). Presented at the International Congress of Phonetic Sciences, Barcelona, Spain.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native

phonological system. *Journal of the Acoustical Society of America*, *109*(2), 775-794.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language

perception of approximants. *Journal of Phonetics*, *20*, 305-350.

Best, C. T., & Tyler, M. (2007). Nonative and second-language speech perception: Commonalities and complementarities. In O. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning : In Honor of James Emil Flege* (pp. 13-34). Philadelphia: John Benjamins Publishing Company.

Bever, T. G. (1975). Cerebral asymmetries in humans are due to the differentiation of two incompatible processes: holistic and analytic. *Annals of the New York Academy of Sciences*, *263*, 251-262.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977-985.

Bradlow, A. R., Pisoni, D. B., Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*(4), 2299-2310.

Brandler, S., & Rammsayer, T. H. (2003). Differences in Mental Abilities between Musicians and Non-Musicians. *Psychology of Music*, *31*(2), 123-138.

Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Language and Speech*, *49*(4), 451-493.

Broselow, E., Hurtig, R. R., & Ringen, C. (1987). The perception of second language prosody. In S. Weinberger & G. Ioup (Eds.), *Interlanguage Phonology: The Acquisition of a Second Language Sound System* (pp. 350-361). Cambridge, MA: Newbury House Publishers.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., & Keller, P. (1996). Perception of Lexical Tone Across Languages: Evidence for a Linguistic Mode of Processing. In *Proceedings of the 4th ICSLP* (pp. 2514-2517). Presented at the 4th International Conference on Spoken Language Processing, Philadelphia, PA.

Burnham, D., & Mattock, K. (2007). The perception of tones and phones. In O. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* (pp. 259-279). Philadelphia: John Benjamins Publishing Company.

Chan, A. S., Ho, Y., & Cheung, M. (1998). Music training improves verbal memory. *Nature*, *396*, 128.

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

Chao, Y. (1930). A system of tone letters. *Maitre Phonétique*, *45*, 24-27.

Chun, D. M., & Plass, J. L. (1996). Effects of Multimedia Annotations on Vocabulary Acquisition. *The Modern Language Journal*, *80*(2), 183-198.

Curtin, S., Goad, H., & Pater, J. V. (1998). Phonological transfer and levels of representation: the perceptual acquisition of Thai voice and aspiration by English

and French speakers. *Second Language Research*, *14*(4), 389-405.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language & Speech*, *29*(3), 201-220.

Cutler, A., & Chen, H. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, *59*(2), 165-179.

Cutler, A., & Clifton, C. E. (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 183-196). Hillsdale, NJ: Erlbaum.

Davidson, L., Shaw, J., & Adams, T. (2007). The effect of word learning on the perception of non-native consonant sequences. *Journal of the Acoustical Society of America*, *122*(6), 3697-3709.

Delogu, F., Lampis, G., & Belardinelli, M. O. (2006). Music-to-language transfer effect: may melodic ability improve learning of tonal languages by native nontonal speakers? *Cognitive Process*, *7*, 203-207.

Delogu, F., Lampis, G., & Belardinelli, M. O. (2009). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *European Journal of Cognitive Psychology*. Retrieved from http://www.informaworld.com.proxy.lib.sfu.ca/10.1080/09541440802708136

Ehrman, M., & Leaver, B. L. (2003). Cognitive styles in the service of language learning. *System*, *31*(3), 393-415.

Ellis, N., & Beaton, A. (1993a). Factors Affecting the Learning of Foreign Language Vocabulary: Imagery Keyword Mediators and Phonological Short-term Memory. *The Quarterly Journal of Experimental Psychology*, *46A*(3), 533-558.

Ellis, N. C., & Beaton, A. (1993b). Psycholinguistic Determinants of Foreign Language Vocabulary Learning. *Language Learning*, *43*(4), 559-617. doi:10.1111/j.1467-1770.1993.tb00627.x

Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics*, *24*, 369-383.

Flege, J. E. (1988). The production and perception of speech sounds in a foreign languages. In H. Winitz (Ed.), *Human Communication and Its Disorders, A Review* (pp. 224-401). Norwood, N.J.: Ablex.

Flege, J. E. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233– 277). Baltimore: York Press.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter.

Flege, J. E., Bohn, O., & Jang, S. (1997). The effect of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics*, *25*, 437-470.

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, *25*, 169-186.

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America*, *95*(6), 3623-3641.

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

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 349-366.

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*, 149-175.

Gandour, J. T., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, G. (2000). A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience*, *12*(1), 207-222.

Gandour, J. T., Wong, D., Lowe, M., & Dzemidzic, M. (2002). A Cross-Linguistic fMRI Study of Spectral and Temporal Cues Underlying Phonological Processing. *Journal of Cognitive Neuroscience*, *14*(7), 1076-1087.

Gardner, R. C., Lalonde, R. N., & Moorcroft, R. (1985). The Role of Attitudes and Motivation in Second Language Learning: Correlational and Experimental Considerations. *Language Learning*, *35*(2), 207-227.

Gordon, E. (1965). *Music Aptitude Profile*. Boston: Houghton Mifflin.

Gordon, E. (1989). *Advanced Measures of Music Audiation*. Gia Publications.

Gottfried, T. L. (2007). Music and language learning: effects of musical training on learning L2 speech contrasts. In O. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* (pp. 221-237). Philadelphia: John Benjamins Publishing Company.

Grabe, E., Rosner, B. S., García-Albea, J. E., & Zhou, X. (2003). Perception of English Intonation by English, Spanish and Chinese Listeners. *Language and Speech*, *46*(4), 375-401.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Research surveys in linguistics. Cambridge: Cambridge University Press.

Halle, P. A., Chang, Y., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*, 395-421.

Hallé, P. A., Best, C. T., & Levitt, A. (1999). Phonetic vs. phonological influences on French listeners: perception of American English approximants. *Journal of Phonetics*, *27*, 281-306.

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context and talker variability. *Applied Psycholinguistics*, *24*, 495-522.

Hardison, D. M. (2005). Second-language spoken word identification: Effects of perceptual
training, visual cues, and phonetic environment. *Applied Psycholinguistics*, *26*, 579-596.

Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, *24*(1), 5-33.

Hirata, Y. (2004). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *Journal of the Acoustical Society of America*, *116*(4), 2384-2394.

Hirata, Y., Whitehurst, E., & Cullings, E. (2007). Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates. *Journal of the Acoustical Society of America*, *121*(6), 3837-3845.

Ingram, J. C. L., & Park, S. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, *25*, 343-370.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57.

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /th/-dh/ contrast by francophones. *Perception & Psychophysics*, *40*(4), 205-215.

Kao, D. L. (1971). *Structure of the syllable in Cantonese*. The Hague: Mouton.

King, R. D. (1967). Functional Load and Sound Change. *Language*, *43*(4), 831-852.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*, *7*(3), 302-307.

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics*, *23*, 81-95.

Krishnan, A., Swaminathan, J., & Gandour, J. T. (2008). Experience-dependent Enhancement of Linguistic Pitch Representation in the Brainstem Is Not Specific to a Speech Context. *Journal of Cognitive Neuroscience*, *21*(6), 1092-1105.

Ladd, D. R. (1978). *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.

Ladd, D. R. (1997). *Intonational Phonology*. Cambridge: Cambridge University Press.

van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal.

Papers in Linguistics: International Journal of Human Communication, *13*(2), 201-277.

Lee, C., & Hung, T. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America*, *124*(5), 3235-3248. doi:10.1121/1.2990713

Lee, Y., Vakoch, D. A., & Wurm, L. H. (1996). Tone Perception in Cantonese and Mandarin: A Cross-Linguistic Comparison. *Journal of Psycholinguistic Research*, *25*(5), 527-542.

Lenneberg, E. (1967). *Biological foundations of language.* New York: Wiley.

Levitin, D. J., & Menon, V. (2003). Musical structure is processed in "language" areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *NeuroImage*, *20*, 2142-2152.

Lively, S. E., Pisoni, D. B., Yamada, R., Tohkura, Y., & Yamada, T. (1994a). Training Japanese listeners to identify English/r/and l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, *96*(4), 2076-2087.

Lively, S. E., Pisoni, D. B., Yamada, R., Tohkura, Y., & Yamada, T. (1994b). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, *96*, 2076-2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874-886.

MacKay, I. R. A., Flege, J. E., Piske, T., & Schirru, C. (2001). Category restructuring during second-language speech acquisition. *Journal of the Acoustical Society of America*, *110*(1), 516-528.

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of Second-Language Acquisition and Perception of Speech in Noise. *J Speech Lang Hear Res*, *40*(3), 686-693.

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, *30*(2), 229-258.

McClaskey, C. L., Pisoni, D. B., & Carrell, T. D. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception & Psychophysics*, *34*(4), 323-330.

Milovanov, R., Pietilä, P., Tervaniemi, M., & Esquef, P. A. (2010). Foreign language pronunciation skills and musical aptitude: A study of Finnish adults with higher education. *Learning and Individual Differences*, *20*(1), 56-60.

Milovanov, R., Tervaniemi, M., & Gustafsson, M. (2004). The impact of musical aptitude in foreign language acquisition (pp. 717-718). Presented at the 8th International Conference on Music Perception & Cognition, Evanston, IL.

Mochizuki, M. (1981). The identification of /p/ and /l/ in natural and synthesized speech. *Journal of Phonetics*, *9*, 283-303.

Morton, D. (1976). *The Traditional Music of Thailand*. Los Angeles: University of California Press.

Nguyên, V. L., & Edmondson, J. (1997). Tones and voice quality in modern northern Vietnamese: Instrumental case studies. *Mon-Khmer Studies*, *28*, 1-18.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7),

674-681.

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

Pater, J., Stager, C. L., & Werker, J. (2004). The perceptual aquisition of phonological contrasts. *Language*, *80*(3), 384-402.

Patkowski, M. S. (1990). Age and Accent in a Second Language: A Reply to James Emil Flege. *Applied Linguistics*, *11*(1), 73-89. doi:10.1093/applin/11.1.73

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nat Neurosci*, *6*(7), 688-691. doi:10.1038/nn1083

Pham, A. H. (2003). *Vietnamese Tone a New Analysis*. Outstanding dissertations in linguistics. New York: Routledge.

Pierrehumbert, J., & Hirschberg, J. (1990). The Meaning of Intonational Contours in the Interpretation of Discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in Communication* (pp. 271-311). Cambridge, MA: MIT Press.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some Effects of Laboratory Training on Identification and Discrimination of Voicing Contrasts in Stop Consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 297-314.

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(5), 976-986.

Pruitt, J. S. (1995). Perceptual training on Hindi dental and retroflex consonants by native English and Japanese speakers. *Journal of the Acoustical Society of America*, *97*, 3417.

Rochet, B. (1995). Perception and production of second-language speech sounds by adults. In *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379-410). Baltimore: York Press.

Schellenberg, E. G. (2003). Does exposure to music have beneficial side effects? In *The Cognitive Neuroscience of Music*. Oxford: Oxford University Press.

Schwanhäuβer, B. (2008). *Lexical tone perception and production : the role of language and musical background* (Doctoral dissertation). University of Western Sydney.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. New York: Newbury House/Harper & Row.

Slevc, L. R., & Miyake, A. (2006). Individual Differences in Second-Language Proficiency: Does Musical Ability Matter? *Psychological Science*, *17*(8), 675-681.

Snodgrass, J. G., & Vanderwart, M. (1980). A Standardized Set of 260 Pictures: Norms for Name Agreement, Image Agreement, Familiarity, and Visual Complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174-215.

So, C. K. (2005). The effect of L1 prosodic backgrounds of Cantonese and Japanese speakers on the perception of Mandarin tones after training. *Journal of the Acoustical Society of America*, *117*(4), 2427.

So, C. K. (2006a). *Effects of L1 Prosodic Background and AV training on Learning Mandarin Tones by Speakers of Cantonese, Japanese, and English* (Doctoral dissertation). Simon Fraser University.

So, C. K. (2006b). Perception of non-native tonal contrasts: Effects of native

phonological and phonetic influences. In P. Warren & C. I. Watson (Eds.), *Proceedings of the 11th Australian International Conference on Speech Science & Technology* (pp. 438-443). Presented at the 11th Australian International Conference on Speech Science & Technology, Auckland, New Zealand.

Song, J. H., Skoe, E., Wong, P. C. M., & Kraus, N. (2008). Plasticity in the Adult Human Auditory Brainstem following Short-term Linguistic Training. *Journal of Cognitive Neuroscience*, *20*(10), 1892-1902.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640), 381-382.

Studdert-Kennedy, M., & Shankweiler, D. (1970). Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, *48*, 579-594.

Surendran, D., & Levow, G. (2004). The Functional Load of Tone in Mandarin is as High as that of Vowels. In *Proceedings of Speech Prosody 2004*. Presented at the Speech Prosody 2004, Nara, Japan.

Szwedek, A. (1986). *A linguistic analysis of sentence stress*. Tübinger Beiträge zur Linguistik (Tub., Vol. 298). Tübingen: Gunter Narr Verlag.

Tamminen, J., & Gaskell, M. G. (2008). Newly learned spoken words show long-term lexical competition effects. *The Quarterly Journal of Experimental Psychology*, *61*(3), 361-371.

Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi, R. J., & Näätänen, R. (2001). Superior Formation of Cortical Memory Traces for Melodic Patterns in Musicians. *Learning & Memory*, *8*(5), 295-300.

Trofimovich, P., & Baker, W. (2006). Learning Second Language Suprasegmentals:

Effect of L2 Experience on Prosody and Fluency Characteristics of L2 Speech. *Studies in Second Language Acquisition*, *28*, 1-30.

Van Lancker, D., & Fromkin, V. A. (1973). Hemispheric specialization for pitch and "tone": Evidence from Thai. *Journal of Phonetics*, *1*, 101-109.

Vance, T. J. (1976). An experimental investigation of tone and intonation in Cantonese. *Phonetica*, *33*, 368-392.

Vance, T. J. (1977). Tonal distinction in Cantonese. *Phonetica*, *34*, 93-107.

Wang, X. (2006). Perception of L2 Tones: L1 Lexical Tone Experience May Not Help. In *Speech Prosody 2006*. Presented at the Speech Prosody 2006, Dresden, Germany.

Wang, Y., Behne, D. M., Jongman, A., & Sereno, J. A. (2004). The role of linguistic experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics*, *25*, 449-466.

Wang, Y., Jongman, A., & Sereno, J. A. (2001). Dichotic perception of Mandarin tones by Chinese and American listeners. *Brain and Language*, *78*, 332-348.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, *113*(2), 1033-1043.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649-3658.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese Listeners to Perceive Thai Tones: A Preliminary Report. *Language Learning*, *54*(4), 681-712.

Wayland, R. P., & Li, B. (2008). Effects of two training procedures in cross-language

perception of tones. *Journal of Phonetics*, *36*, 250-267.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*, 1-25.

Weinberger, S. (1990). Minimal segments in second language phonology. In J. Leather & A. James (Eds.), *New Sounds 90, Proceedings of the 1990 Amsterdam Symposium on the Acquisition of Second-Language Speech*. Presented at the 1990 Amsterdam Symposium on the Acquisition of Second-Language Speech, Amsterdam: University of Amsterdam.

Werker, J., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, *52*, 349-355.

Werker, J., & Tees, R. C. (2002). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, *25*, 121-133.

Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*, 565-585.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*(4), 420-422.

Yamada, R. (1993). Effect of extended training on /p/ and /l/ identification by native speakers of Japanese. *The Journal of the Acoustical Society of America*, *93*, 2391.

Yip, M. (2002). *Tone*. Cambridge, UK: Cambridge University Press.