Hakka tone training for native speakers of tonal and nontonal languages

Jessica L. L. Chin

Supervisor: Dr Mark Antoniou

A thesis submitted in partial fulfilment of the requirements for the degree of

MASTER OF RESEARCH

The MARCS Institute for Brain, Behaviour and Development

Western Sydney University

9 December, 2018

Statement of Authentication

I declare that this thesis has not been submitted, either in full or in part, for a degree at this or any other university or institution. The work presented in this thesis is original research, except as acknowledged in the text. This study was approved by the Western Sydney University Human Research Ethics Committee (approval number HI 0988, funding source 20211/73040).

[redacted]

Jessica L. L. Chin

Table of Contents

Tables

Figures

Abstract

Language learning becomes increasingly difficult when novel linguistic features are introduced. Studies have shown that learners from various language backgrounds can be trained to perceive lexical tone, which assigns meaning to words using variations in pitch. In this thesis, we investigated whether native speakers of tonal Mandarin Chinese and tonal Vietnamese outperformed native speakers of nontonal English when learning Hakka Chinese tones following five sessions of tone training, and whether the complexity (i.e., density) of a listener's native tone inventory facilitated nonnative tone learning. All groups improved in tone identification and tone word learning following training, with improvements persisting three weeks following the cessation of training. Although both tonal groups outperformed the English group in most tasks, the Mandarin group showed the most consistent advantages over the English group across tasks. Findings suggest that tone experience bolsters tone learning, but density of the tone inventory does not provide an advantage. Confusion patterns offer detailed insight of the interaction between nonnative tones and native tonal and intonational categories.

Keywords: Hakka Chinese, tone training, language experience

Chapter 1: Introduction

Compared to the ease of language acquisition in infancy and early childhood, language learning in adulthood is a difficult task. Although infants are initially able to successfully discriminate between nonnative sound contrasts, this ability declines by 6 to 12 months coinciding with the time that infants show improvement in discriminating sounds in their native language (Kuhl et al., 2006). As we age, it becomes difficult to acquire novel sounds and linguistic features differing from those occurring in our native language. There has been extensive research on how adults perceive nonnative segments, that is, consonants (Best, McRoberts, & Sithole, 1988; Guion, Flege, Akahane-Yamada, & Pruitt, 2000) and vowels (Flege, Bohn, & Jang, 1997), as well as how adults learn higher aspects of language such as vocabulary (Paribakht & Wesche, 1996) and syntax (Clahsen & Muysken, 1986). Research has also been conducted on adults' perception of suprasegmentals, that is, the contrastive features that occur beyond the level of segments, such as rhythm and intonation (Ramus & Mehler, 1999). While intonation is the variation of spoken pitch at the postlexical or sentence level (Ladd, 2008), lexical tone is a feature which utilises pitch pattern variations to distinguish the meanings of words (Yip, 2002). Interestingly, children with no tone language experience are better at discriminating nonnative consonant contrasts than nonnative tone contrasts, but their adult counterparts are better at discriminating tones than consonants (Burnham & Mattock, 2007). In this thesis, we investigate how lexical tone is processed and learned by adults from tonal and nontonal language backgrounds. Across five sessions, participants were trained to learn the tones of Hakka Chinese. Performance was compared between nontonal Australian English speakers, tonal Mandarin, and tonal Vietnamese speakers to determine the effect of language background on the learning of a nonnative tone system. Additionally, learners were invited to return three weeks after the cessation of training to complete a tone identification retention test, which observed whether any training-related gains were maintained.

The following sections provide an overview of the literature on tone perception and tone training. Section 1.1 introduces pitch as it is relevant in linguistic contexts, and the characteristics of lexical tone and intonation are outlined. Section 1.2 covers the factors that influence tone perception, namely language and musical background. Section 1.3 compares the speech perception models which have been used to account for patterns of nonnative tone perception, although they were not originally developed for this purpose. In Section 1.4, findings from past tone training studies are critically reviewed and predictors of successful tone learning are discussed. Section 1.5 presents the research questions and hypotheses for the study, along with information on the tonal and intonational systems of the training language and the learner groups' languages. Participant demographics, material creation and the training protocol are explained in Chapter 2. The results of each task are presented in Chapter 3. In Chapter 4, tone confusion patterns are compared between groups and are examined using one of the aforementioned speech perception models. The limitations of this study are evaluated and future studies are proposed. Finally, Chapter 5 summarises the findings of this study and its implications for nonnative models of lexical tone perception.

1.1. Pitch

Pitch refers to the perceptual correlate of the fundamental frequency (F_0). In a sound signal, the number of pulses per second is the F_0 and is measured in Hertz (Hz), referring to a cycle per second (Yip, 2002). Both music and language exploit time-varying pitch patterns to convey information. In the music domain, pitch is considered a perceptual property attributed to musical melodies whereby pitch variations are detected by the rate of vibration of a sound (Plack, Oxenham, Fay, & Popper, 2006). In the language domain, pitch serves numerous purposes: lexical tone languages may use pitch to distinguish the meanings of words (see Section 1.1.1); and pitch may appear at the sentence level in the form of intonation (see Section 1.1.2).

1.1.1 Lexical tone

Lexical tone is a phonologically contrastive feature that occurs in over half of the world's languages. In these languages, a syllable or word, such as *ma* in Mandarin Chinese, will have different meanings depending on the pitch contour assigned to the syllable: for instance, *ma* produced with a high level tone means 'mother', while *ma* produced with a mid-to-high rising tone means 'hemp' (M. Y. Chen, 2000). Geographically, languages which utilise tone appear most prominently in East and Southeast Asia, Sub-Saharan Africa, the Pacific, and Central America. Some of the most prevalent language families with tonal characteristics include the Sino-Tibetan, Tai-Kadai and Austroasiatic families in East and Southeast Asia, Niger-Congo in Africa, and Oto-Manguean in Central America (Yip, 2002). The tone languages relevant to this thesis include Vietnamese, a Mon-Khmer language in the Austroasiatic language family, and the Hakka and Mandarin dialects of Chinese, a branch of the Sino-Tibetan language family (see Sections 1.5.1-1.5.3 for descriptions of these tone systems). It should be noted that not all languages with tone as a contrastive feature can be considered tone languages; while pitch-accented languages do not have tone systems, pitch in these languages can be lexically contrastive when it occurs in an accented syllable in a word (Beckman, 2012). Japanese, Swedish, Lithuanian and Serbo-Croatian are examples of pitch-accented languages (Yip, 2002).

Tone languages differ in the number of tones in their tone inventory, and individual tones can differ in pitch height, as well as the direction and trajectory of the pitch contour. Tones are usually classified as either level or contour. Although contour tones with more than one direction (for example, a rising-falling tone) are sometimes labelled complex tones (Yip, 2002), use of such descriptors varies across the literature. Level, or static, tones are characterised by a steady pitch contour that does not change in direction but may vary in pitch height. Contour, or dynamic, tones are characterised by both an abrupt movement of the direction and trajectory of a pitch slope; unlike level tones, their pitch can rise and fall along the syllable (Abramson, 1978). Though often categorised as a contour tone, a complex tone can fall *and then* rise in a concave or ‘dipping’ contour, or rise and then fall in a convex contour. Concave tones appear more often than convex tones, with complex tones appearing most commonly in Asian languages (Yip, 2002). Another classification of tone present in some tone languages is the checked tone, which refers to tones articulated in ‘checked’ syllables that end in a stop consonant (or glottal stop). In the Cantonese and Hakka dialects of Chinese, these checked syllables end with the unreleased voiceless plosives /p̚/, /t̚/, and /k̚/ (Bauer & Benedict, 1997; Hashimoto, 2010). Since these checked syllables are shorter in duration than sonorant-final syllables, level tones usually occur in these contexts (Yip, 2002), though not always; Hakka has both high level and high falling checked tones (Hashimoto, 2010). In other languages such as Vietnamese, the checked tones are allophones of the unchecked tones (Brunelle, Nguyên, & Nguyên, 2010). Additional contrastive elements such as phonation and vowel quality are present in some tone systems. In Northern Vietnamese, some tones are produced with laryngealisation and glottalisation; however, these features play a limited role in the Central and Southern dialects (Brunelle, 2009; Brunelle et al., 2010; Nguyễn & Edmondson, 1997).

Three generalisations can be made regarding the types of tones present within a tone system: a system with contour tones will have level tones; a system with complex tones will have simple contour tones; and a system with rising tones will have falling tones (Zhang, 2004). Also relevant to tone systems is tone sandhi, which is a phonological change of a word or morpheme's tone depending on what words or morphemes occur adjacent to it. In Mandarin, a low-dipping tone changes to a mid-rising tone when the following tone is also a low-dipping tone, that is, the first dipping tone changes while the second dipping tone remains unchanged (M. Y. Chen, 2000). Discussions of tone sandhi in this thesis will be limited to this section as all stimuli presented to the listeners are monosyllabic.

There are multiple forms of tone notation that linguists have developed to visualise the pitch contours of a tone in orthography. African tone languages use accent marks to characterise different tones, while Central American tone languages use numbers to convey the height and contour of a tone. This thesis utilises the Asian notation style for tone languages, which was developed as 'tone letters' and numerical tone values similar to those used in Central American tone notation (Chao, 1930). Unlike the Central American notation style, where a value of 1 represents the highest pitch and 5 the lowest, Chao's numerical tone notation categorises 1 as the lowest pitch value and 5 as the highest. To show the movement of a tone's pitch contour, each pitch target is assigned a value. A level tone with two pitch targets would have two identical values; for example, /ma/ produced with a high level tone would be transcribed as /ma^{55}/. A contour tone would show the rise or fall of the contour, such as in the mid-to-high rising tone /ma^{35}/. In the case of a complex contour tone, each of the three pitch targets would be notated, such that a dipping tone would be transcribed as /ma^{214}/. Similarly, tone letters provide a more visual depiction of a pitch contour. Beginning on the left at its initial pitch height, the letter would resemble the pitch contour's movement along the syllable, finally reaching a vertical line that serves as an end point for the contour. Using tone letters, the previous examples would be transcribed as /ma˥/, /ma˧˥/, and /ma˨˩˦/. Alternatively, linguists have assigned citation tone numbers for tone systems. For instance, Mandarin's first tone (T1) always refers to the high level tone 55, and can also be transcribed as /ma1/. For purposes of readability, the thesis will use Chao's numerical notation to convey tones, as well as the citation tone number for the language.

With the diverse array of tone systems amongst even languages from the same language family, there have been attempts to categorise systems in terms of complexity. However, the definitions of complexity in the literature are both inconsistent and insufficient. The most prominent definition of tonal or prosodic complexity is the number of tones within a tone system, such that a tone language with a denser tone system is considered more complex than one with fewer tones (Maddieson, Bhattacharya, Smith, & Croft, 2011; Tong & Tang, 2016). Maddieson et al. (2011) categorises tone systems as simple (two tones), moderately complex (three tones), or complex (four or more tones). This definition may not encompass the other characteristics of tone detailed in preceding paragraphs, such as the type of tones present within a tone system or an individual tone's relation to other tones. There have been more holistic accounts of complexity, albeit in relation to individual tones rather than the whole system. These approaches have defined complexity as the number of pitch targets within an individual tone (two for level and simple contour tones and three for complex contour tones), the excursion or trajectory of the pitch contour between the targets, and the direction of the contour (Zhang, 2004). With all other aspects of tone considered, there are potential issues in determining a tone system's complexity by density alone. Few studies in the literature have compared nonnative tone perception between different tone language groups, which makes it difficult to evaluate the effectiveness of this definition (see Sections 1.2.1 and 1.4.1 for a review of these studies). However, creating a new definition for tonal complexity is beyond the scope of this thesis. Rather, this thesis will test the predictive power of the most commonly used definition of complexity on nonnative tone learning. As the results will show, establishing a modified definition for the complexity of an entire tone system is a topic worth investigating in future research.

1.1.2 Intonation

As stated previously, intonation involves the variation of suprasegmental features, namely pitch, at the postlexical level. Unlike lexical tone, intonation is present in both tone and nontone languages and does not distinguish the meanings of segmentally-identical syllables or words. Instead, it can provide discoursal meaning to an utterance, such as prompting an answer from an interlocutor, or attitudinal meaning, such as condescension (Cruttenden, 1997). A notable system used for transcribing intonation is the Tones and Breaks Indices (ToBI) system, which consists of four notation tiers: orthography, tone, break-index and miscellaneous (Beckman, 2012; Beckman & Hirschberg, 1994). The orthographic tier contains the orthographic transcription for an utterance, while the break-index tier evaluates the degree of juncture between words, including at the end of an utterance. The miscellaneous tier includes additional information in the utterance, such as silences and laughter. In the tone tier, tones are labelled ‘H’ for high and ‘L’ for low, and are separated into phrasal tones and pitch accents. While phrasal tones occur at the intermediate and intonational phrase boundaries, pitch accents occur at accented syllables. One of the learner groups in this study is the Australian English group, which is a nontone language with an established set of intonational patterns. The intonational system of (Australian) English is outlined in Section 1.5.4.

1.2. Tone perception

Nonnative tone perception is influenced by a range of factors, and these can have facilitating or debilitating effects. While this thesis primarily explores the effects of language experience on tone perception, this section also briefly overviews the effects of musicianship, cognitive ability, and hearing and musical disorders.

1.2.1 Language experience

Studies have shown that native experience in a tone language can modulate tone perception. One particular study revealed that success in distinguishing novel tone minimal pairs was most strongly attributed to proficiency in an East Asian language, including tonal languages (such as Vietnamese, Thai, or a Chinese dialect) as well as nontonal Korean and pitch-accented Japanese (Caldwell-Harris, Lancaster, Ladd, Dediu, & Christiansen, 2015). Native tone language listeners show a clear perceptual advantage over native nontonal language listeners when identifying and discriminating native tones (Gottfried & Suiter, 1997). Though nontonal listeners do show sensitivity to nonnative tone contrasts, only tonal listeners show strong or quasi-categorical perception of speech and nonspeech tone stimuli (Hallé, Chang, & Best, 2004; Wu & Lin, 2008; Xu, Gandour, & Francis, 2006). Nontonal listeners have a tendency to attend to pitch movement that is relevant in their native language, such as pitch signalling postlexical information in Dutch (Braun & Johnson, 2011), and are able to assimilate nonnative tones to their native intonational categories (So & Best, 2014).

Nonnative (L2) experience in a tone language has been shown to benefit listeners regardless of whether their native language (L1) is tonal or nontonal. Native speakers of English with experience in Thai were better at discriminating Thai tones than naïve English speakers, and also showed similar discrimination patterns to native Thai speakers in closed syllables (Wayland & Guion, 2003). Similarly, native Mandarin speakers with experience in Thai and native Thai speakers with experience in Mandarin showed sensitivity to both acoustic and phonological features of the L2 when categorising tones, whereas naïve Mandarin and Thai listeners assimilated the nonnative tones to native categories based on acoustic features such as F_0 height and contour (Wu, Munro, & Wang, 2014). Tonal L2 experience has also been shown to influence the perception of tones in a third language (L3), with native English learners of Mandarin showing different patterns of Cantonese tone discrimination from English monolinguals and native Mandarin speakers (Qin & Jongman, 2016).

A listener's native language also affects how one attends to nonnative tones. Past studies have used multidimensional scaling analyses to determine how tonal and nontonal listeners weight the perceptual dimensions relevant to the processing of tone. Native speakers of a nontonal language such as English, as well as native speakers of pitch-accented languages such as Japanese, attend more to pitch height when perceiving tones, though Japanese speakers also attend to high level tones since high tones mark accented syllables in their native language (Guion & Pederson, 2007). Conversely, native speakers of a tone language attend more to pitch direction and slope (J. T. Gandour & Harshman, 1978; Xu et al., 2006). Despite the pitch direction dimension being underweighted by nontonal language speakers, their nonnative tone categorisation can be improved if provided with explicit instruction to attend to pitch direction and ignore pitch height (Chandrasekaran, Yi, Smayda, & Maddox, 2016). The differences between tone perception by tonal and nontonal language speakers are further supported by evidence showing how tone is processed in the brain. Tone language speakers show a listening advantage in the right ear and activation in the left hemisphere when listening to tones in their native language, but nontonal language speakers process tones bilaterally. Similar to nontonal language speakers, native tone language speakers who are naïve listeners of the target tones do not show significant left hemisphere activation (J. Gandour et al., 2000; Y. Wang, Jongman, & Sereno, 2001). However, other measures, including the mismatch negativity (MMN) in the cortex and the frequency following response (FFR) in the auditory brainstem, show that native tone language experience not only enhances preattentive pitch processing, but also provides benefits to pitch processing in a nonnative tone system (Chandrasekaran, Krishnan, & Gandour, 2009b; Krishnan, Gandour, & Bidelman, 2010b).

The effects of native tone language experience on the perception of nonnative tones have varied across studies. Some studies show that native tone language speakers have an advantage in tone perception over nontonal language speakers (Burnham, Kasisopa, et al., 2015; Schaefer & Darcy, 2014) and nonnative tone language speakers (C.-Y. Lee, Tao, & Bond, 2010), and this advantage extends to multiple-talker contexts (Y. S. Chang, Yao, & Huang, 2017). The benefits of a complex (i.e., denser) native tone system are less clear. Cantonese–Mandarin bilingual children develop phonological awareness earlier than Mandarin monolinguals, and one explanation for their increased tone awareness is their knowledge of Cantonese (X. Chen et al., 2004). Similarly, adult native speakers of Cantonese can distinguish pitch differences easier than Mandarin speakers, and this has been attributed to their denser native tone system (Zheng, Minett, Peng, & Wang, 2012). Although a complex native tone language may facilitate an individual’s sensitivity to pitch, other studies have shown that native tone categories can interfere with nonnative tone perception. No differences were observed between Mandarin and English listeners in the discrimination of nonspeech tones (Bent, Bradlow, & Wright, 2006), with both listener groups exhibiting difficulties in discriminating different tone pairs (Qin & Mok, 2013). While Cantonese and Japanese listeners outperformed nontonal English listeners in Mandarin tone identification, the Cantonese group did not differ from the Japanese group (So & Best, 2010). In a range of studies, the Cantonese listeners’ confusion patterns were strongly suggested to have been caused by interference from their native tone system (Hao, 2012; Tsukada, Xu, & Rattanasone, 2015). In another study, Taiwanese listeners performed significantly worse than Vietnamese and German listeners when discriminating the four level tones in Toura, a Niger-Congo language. It was suggested that the two level tones in Taiwanese (compared to the one level tone in Vietnamese and no tones in German) acted as perceptual magnets and interfered with the discrimination of the nonnative tones (Chiao, Kabak, & Braun, 2011). Additionally, language dominance in bilinguals can affect tone perception: native tone processing in Mandarin–English bilingual listeners was hindered when English was the listener's dominant language, even if Mandarin was acquired first (Quam & Creel, 2017).

The conflicting evidence concerning the effects of native tone language background shows that we have not been able to confirm whether prior tone experience facilitates or hinders the perception of nonnative tones. Perhaps it is more insightful to examine which specific tones different language groups find easier or more difficult to learn, and whether these patterns interact with the listener's native tonal or intonational system. This is often done by utilising speech perception models to make predictions for how listeners might perceive tones in relation to their native categories (see Section 1.3). Also requiring consideration are the aspects of tone perception which are universal across tone languages. Several findings were made by Burnham et al. (2015): listeners from all language backgrounds were able to use visual information to perceive Thai tones, but nontonal English listeners were better at using this information when presented with visual-only stimuli, which might have been due to the underuse of visual information by tonal and pitch-accent language speakers. The authors also found that the DynamicDynamic (rising–falling) tone contrast was the easiest to perceive, and that tone pairs containing a rising tone were also easier to perceive than other tone contrasts. This is supported by evidence that rising tones are more perceptually salient than falling tones, as seen in other behavioural research and FFR studies (Krishnan, Gandour, & Bidelman, 2010a; Krishnan, Xu, Gandour, & Cariani, 2004; Wayland, Zhu, & Kaan, 2015). Based on this review of the literature, predictions can be formulated concerning which tone contrasts listeners will find easiest to perceive. By investigating the relationships between the target language and the learners' native language(s), it is possible to make language-specific predictions for tone perception.

1.2.2 Musical background

Pitch plays a pivotal role in the production and perception of music, with numerous studies showing evidence that musical experience facilitates nonnative tone perception. Nontone language speakers with musical experience outperform their nonmusician counterparts in tone discrimination and identification tasks (C.-Y. Lee & Hung, 2008), and musicians with absolute pitch outperform musicians without (Burnham, Brooker, & Reid, 2015). Musicians show more robust encoding of linguistic pitch in the auditory brainstem, with pitch tracking performance significantly correlating to the age of onset and duration of musical training (Wong, Skoe, Russo, Dees, & Kraus, 2007). Experience in either musical or linguistic pitch provides domain-general benefits at preattentive stages of processing, as both nontone-language-speaking musicians and native tonal-language-speaking nonmusicians show larger MMN responses than nontone-language-speaking nonmusicians. Although MMN responses are largest in the tone group when discriminating a within-category tone contrast, the tone group's accuracy is also the lowest in this condition (Chandrasekaran, Krishnan, & Gandour, 2009a). In another study, musicians showed enhanced MMN when presented with both speech and music sounds compared to tone language speakers and nonmusicians, and were also superior at timbral discrimination (Hutka, Bidelman, & Moreno, 2015). Speech- and music-specific features of pitch processing do not always transfer seamlessly to the other domain. Nontone-language-speaking musicians tend to perform similarly to native tone language speakers in tonal processing, but musical training does not facilitate phonological processing (Delogu, Lampis,

& Belardinelli, 2010). Experience with musical pitch bolsters pitch processing, but the processing of linguistic pitch also requires the ability to make categorical representations; on the other hand, tone language experience can interfere with melodic tone processing, especially when processing pitch contours that fall within a native tone category (D. Chang, Hedberg, & Wang, 2016).

1.2.3 Cognitive ability

Although cognitive ability has been shown to affect tone perception performance, the effect is not as strong as other factors such as musicality (which includes musical aptitude and experience) and linguistic pitch processing ability (Bowles, Chang, & Karuzis, 2016). Cognitive ability appears to interact with tone and music experience. Native tone language listeners and nontone-language-speaking musicians outperformed nontone-language-speaking nonmusicians in a tonal working memory task, with the musicians also showing an advantage in nonauditory working memory over the tone language listeners and nonmusicians (Bidelman, Hutka, & Moreno, 2013). In another study using Mandarin tones, the increase in stimulus complexity from nonspeech to speech stimuli impacted on short-term categorical memory in English listeners. Since short-term memory involves real-time sensory encoding, the increase in stimulus complexity led to a decrease in pitch sensitivity, thus lowering between- and within-category discrimination accuracy. Contrastingly, the Mandarin listeners were not inhibited by stimulus complexity, as categorical representations of their native tones were stored in their long-term memory (Xu et al., 2006).

1.2.4 Hearing and musical disorders

Cochlear implant recipients have been shown to struggle with tone processing in comparison to their normal-hearing counterparts. For instance, native Mandarin listeners with cochlear implants performed significantly worse in Mandarin tone recognition than simulations of normal-hearing listeners, and they failed to sufficiently extract and encode temporal and spectral cues that are crucial for tone recognition (Wei, Cao, & Zeng, 2004). Several training programs have been effective in improving tone perception in cochlear implant recipients. The Computer-Assisted Speech Training program exposed listeners to Mandarin vowel, consonant and tone contrasts, with overall speech perception improving following training (Fu & Galvin, 2007). Cochlear implant recipients improved in question/statement intonation recognition following melodic contour training (Lo, McMahon, Looi, & Thompson, 2015), and improved in musical pitch perception following music training (Gfeller, Guthe, Driscoll, & Brown, 2015).

Sufferers of congenital amusia (also known as tone-deafness) have an impaired ability to perceive pitch, which has debilitating effects on music and tone perception (Ayotte, Peretz, & Hyde, 2002; Peretz, 2001). Amusic speakers of nontone languages generally perform worse in linguistic and musical pitch recognition than controls (Jones, Lucker, Zalewski, Brewer, & Drayna, 2009; F. Liu, Patel, Fourcin, & Stewart, 2010; Nguyen, Tillmann, Gosselin, & Peretz, 2009), with amusics compensating for their pitch processing impairment by attending to the sound duration and intensity of stimuli (Tillmann et al., 2011). Native tone language speakers with amusia also perform worse in tone identification and discrimination than controls, but their deficits do not extend to native tone production (Nan, Sun, & Peretz, 2010) or in conditions where they can rely on other acoustic cues to discriminate stimuli (F. Liu et al., 2012). Providing pitch direction training to amusic Mandarin listeners has been shown to improve their sensitivity to pitch in speech and music, with their performance matching the control group's (F. Liu, Jiang, Francart, Chan, & Wong, 2017).

1.3. Nonnative speech perception models

Several models in the literature have sought to explain the perception of nonnative speech, though their primary focus has been on consonant and vowel contrasts rather than lexical tones. A speech perception model based purely on tone perception has yet to be proposed, but the principles in some segmental speech perception models have been extended to account for nonnative tone perception, if imperfectly. Some of these models include the Speech Learning Model (SLM), the Processing Rich Information from Multidimensional Interactive Representations (PRIMIR) framework, and the Perceptual Assimilation Model (PAM). Brief overviews of these models are provided below. The hypotheses in this thesis relate closest to principles of the Perceptual Assimilation Model, which addresses speech perception by naïve listeners (as well as L2 learners).

1.3.1 Speech Learning Model (SLM)

SLM (Flege, 1995, 2003) explores the changes in the perception and production of L2 sounds (initially segments) across the human lifespan, with a focus on more experienced L2 learners rather than beginners. SLM posits that the mechanisms and processes involved in L1 acquisition are retained in adulthood and can be utilised in L2 learning. Phonetic categories are representations of speech sounds that, when acquired in childhood, are stored in long-term memory. These phonetic categories adapt over a bilingual's lifespan as representations of L1 and L2 phones, all of which share a common phonological space. SLM also comprises of a range of hypotheses. L1 and L2 sounds are more perceptually related to each other as position-sensitive allophones than abstract phonemes. A new L2 phonetic category can be formed if the bilingual can distinguish between the L2 sound and the closest L1 sound. The differences between these two sounds are more likely to be distinguished if the sounds are more phonetically dissimilar to each other, but this likelihood decreases as the age of learning the L2 increases. Diaphones, which consist of L1 and L2 sounds that could not be formed into separate phonetic categories, are represented by a single phonetic category in perception and production. A bilingual's phonetic category for an L2 sound can differ from a monolingual's category for the same sound for the following reasons: the bilingual's L2 category is shifted away from the closest L1 category to better distinguish the two categories, or the bilingual's weighting of linguistic features differs from the monolingual's. Lastly, an L2 sound will eventually be produced by the bilingual in the way it is represented as a phonetic category.

Some studies have examined the perception of L2 tone categories through the lens of the SLM framework. As mentioned in Section 1.2.1, native speakers of tonal languages attend to the pitch direction and slope of a tone (J. T. Gandour & Harshman, 1978; Xu et al., 2006). Native speakers of nontonal English and pitch-accented Japanese attended more to pitch height when listening to Mandarin tones, though the Japanese group also attended to high level tones, which are present in Japanese accented syllables (Guion & Pederson, 2007). Native English speakers who were experienced L2 Mandarin learners, however, had learned to use pitch slope information to perceive tones. In line with SLM's hypotheses, these results indicate that second language experience allows bilinguals to reweight acoustic cues in a way that differs from monolingual listeners. Another study that compared Mandarin–Thai and Thai–Mandarin bilinguals to Mandarin and Thai listeners with no other tone experience found that both bilingual groups were sensitive to the phonetic and phonemic similarities between tones in their L1 and L2, but naïve listeners were sensitive to phonetic similarities only (Wu et al., 2014).

The above studies demonstrate that the postulates and hypotheses of SLM are transferable to research involving L2 tone perception.

1.3.2 Processing Rich Information from Multidimensional Interactive Representations (PRIMIR)

Unlike the other two speech perception models outlined in this section which address speech perception in adults, PRIMIR (Werker & Curtin, 2005) was developed in response to conflicting findings in infant speech perception research. While some infants were able to use phonetic and phonological distinctions when learning words, other infants could not. PRIMIR accounts for these discrepancies in the literature whilst making predictions concerning how infants perceive speech and acquire language. According to this framework, infants are able to access rich information in speech and can organise this information into the general perceptual, word form, and phoneme planes. The rich information is sorted into these three representational planes using three dynamic filters. Firstly, the infant's initial biases toward speech (particularly infant-directed speech), rhythm in speech, and other language features help the infant attend to the speech signal. Then, the infant's developmental level and the language learning task, such as sound discrimination or word learning, filters which information the infant attends to, stores and utilises. The linguistic signal itself contains features which may be more salient than others, whether it is acoustic, multisensory or contextual.

PRIMIR was initially centred on the processing of segmental contrasts, as well as the processing of talker, affect, and other indexical cues, though there has been recent discussion on its applicability to tone perception (Curtin & Werker, 2018). Nonnative tone discrimination in infants first shows a decline in accuracy at 6 to 9 months, followed by a rebound at 18 months

(L. Liu & Kager, 2014). However, a later study showed that monolingual and bilingual infants were not able to use tonal information to learn words at 18 months (L. Liu & Kager, 2018). It is posited that this is due to the perception of nonnative contrasts being restricted to words at 18 months, which results in infants perceiving tones as phonemic features instead (Curtin & Werker, 2018). The extension of PRIMIR to nonnative tone processing is capable of generating more robust predictions to drive future infant speech perception research.

1.3.3 Perceptual Assimilation Model (PAM)

PAM (Best, 1995) was initially developed to address how naïve listeners categorise nonnative speech segments in relation to their native phonetic categories. Since then, the principles in PAM have been applied to speech perception in L2 learners (Best & Tyler, 2007) and infants (Best, Tyler, Goldstein, & Nam, 2016), and have also been used to generate predictions for nonnative tone perception studies (Best, forthcoming; Reid et al., 2015; So & Best, 2010, 2014). While SLM focuses on the difference between an L2 sound and the closest L1 category to determine the formation of a new category, PAM examines how the nonnative sound will assimilate to the closest L1 category. There are three main patterns of perceptual assimilation for nonnative sounds. A Categorised nonnative speech sound is one that is heard as either a good, mediocre or poor exemplar of a native category. An Uncategorised speech sound is not heard as a single native category and may fall between native categories. A Non-Assimilated speech sound is heard as a nonspeech sound and is therefore not assimilated into the native phonological space. Predictions can also be made on a listener's ability to discriminate between a nonnative sound contrast by observing the perceptual assimilation pattern for the two segments. A nonnative contrast shows Two Category (TC) assimilation when each sound is assimilated into a different native category, and excellent discrimination between the sounds is predicted. Conversely, when two nonnative sounds are heard as equally good or poor exemplars of the same native category, the contrast falls within the Single Category (SC) assimilation type and poor discrimination is expected. The Category Goodness (CG) assimilation type occurs when two nonnative sounds are assimilated into the same native category, but one of the sounds is more acceptable as a native sound than the other. Discrimination of a nonnative contrast in this pattern is predicted to range between moderate to very good. Uncategorised-Categorised (UC) assimilation occurs when both nonnative sounds fall within the native phonological space, but one sound is assimilated into a native category and the other is not. Very good discrimination is expected. A nonnative contrast falls into the Uncategorised-Uncategorised (UU) assimilation type when both sounds cannot be assimilated into a native category. The contrast is predicted to be discriminated moderately well or poorly, depending how closely they occur to each other and to native categories. Lastly, when both nonnative sounds are heard as nonspeech sounds, the contrast is considered Non-Assimilable (NA) and good to very good discrimination is predicted. While rare, the NA pattern has been studied: when adult American English listeners and infants were exposed to Zulu click contrasts, the adults heard the clicks as nonspeech and discriminated the contrasts with high accuracy. Infants also continued to discriminate the click contrasts consistently at 12 to 14 months of age (Best et al., 1988).

PAM has been used as a framework to understand language-specific patterns of assimilation in nonnative tones. So and Best (2010) presented Mandarin tone stimuli to three naïve listener groups—tonal Hong Kong Cantonese, pitch-accented Japanese, and nontonal Canadian English. They hypothesised that nonnative tone perception was influenced by the listeners' L1 prosodic system (i.e., tone, pitch accent or intonation) rather than the degree of tone language experience. Results showed that the Cantonese and Japanese groups outperformed the English listeners, though the Cantonese listeners did not outperform the Japanese listeners. Tone pairs that shared phonetic similarities in pitch contour and pitch height at either onset or offset were more confusable by all groups than tone pairs that were phonetically dissimilar. As predicted, Cantonese listeners struggled to perceive the T1–T4 (55–51) contrast because both tones are similar to the allotones of Cantonese T1 (55), and T2–T3 (35–214) because they both show similarities to Cantonese T2 (25). Although the Japanese listeners were expected to assimilate Mandarin T2 (35) and T4 (51) to their native LH and HL pitch accent patterns, they actually struggled to identify these tones in the study, possibly because they were not yet able to map the nonnative tones to their L1 pitch accent patterns. Several assimilation types were proposed for the English listeners' perception of tone contrasts, in line with suggestions from an earlier study testing Mandarin tone perception by Taiwan Mandarin and French listeners (Hallé et al., 2004). A nontonal listener may find nonnative tones Uncategorisable, since pitch contours occur in nontonal languages (such as English and French) as intonation at the sentential level; Non-Assimilable, as the listener may hear the tones as nonspeech melodic contours; or Categorisable if the nonnative tones are assimilated into the listener's native prosodic categories (such as Mandarin T2 [35] to the rising question intonation in English or stunned intonation in French).

Later studies further examined the influence of native intonational categories on the perception of nonnative tones. In one study, French and English listeners heard nonnative Mandarin tones presented in a sentence environment and mapped them to their native intonational categories (So & Best, 2014). Unlike English, French does not use lexical stress, but the French listeners showed higher discrimination accuracy for the Mandarin tones than the English listeners. This suggests that the French listeners were more sensitive to the phonetic features of the tones, while the English listeners considered both lexical stress and intonation when hearing the tones. Reflecting past findings (So & Best, 2010), both listener groups had the most difficulty discriminating the T1–T4 and T2–T3 contrasts. Though the listeners were predicted to hear Mandarin T1 (55) and T3 (214) as the *Statement* intonation pattern, T2 (35) as *Question*, and

T4 (51) as *Exclamation*, these predictions were not upheld by the tone discrimination results. This was possibly attributed to overlaps in category choices, tonal coarticulation effects, and weak categorisations of tones. Hao (2012) tested L1 English and L1 Cantonese learners of Mandarin on the perception and production of Mandarin tones, and both groups performed similarly in terms of accuracy. Universal difficulty of certain tone contrasts was cited as an additional challenge listeners may face when perceiving L2 tones, alongside difficulties caused by the interaction between L2 sounds and L1 categories. The T2–T3 contrast in Mandarin was once again the most difficult to discriminate for both listener groups, and the Cantonese listeners also showed difficulties in distinguishing the T1–T4 contrast.

In one study, no predictions could be tested for the English listeners, as they showed weak categorisations (<50% for each category) for Thai tones (Reid et al., 2015). The Mandarin group showed stronger categorisations (which aligned more closely to PAM predictions) than the Cantonese group, perhaps due to Cantonese having more tones and category options. Both Mandarin and Cantonese groups mapped Thai falling tones to L1 level tones, and Thai rising tones to L1 rising tones. Wu et al. (2014) examined L2 tone perception by L1 Mandarin and L1 Thai listeners with and without experience in the other tone language. While inexperienced listeners mostly assimilated the nonnative tones to the most phonetically similar L1 category, the experienced listeners also showed assimilation patterns that take into account the allotonic variation of Mandarin T3 (214) as T2 (35) and a low falling tone. Experienced learners of Mandarin assimilated the Thai low falling tone (21) to Mandarin T3 (214) and the Thai low falling rising tone (214) to Mandarin T2 (35), and experienced learners of Thai showed the same assimilation patterns for Mandarin.

In a study training Mandarin and English listeners to learn Cantonese tones, Mandarin listeners were able to accurately identify some nonnative tones which corresponded closely to their native categories: Cantonese T1 (55) was mapped to Mandarin T1 (55), Cantonese T2 (25) to Mandarin T2 (35), and Cantonese T3 (33) to Mandarin T3 (214) or the Mandarin neutral tone (Francis, Ciocca, Ma, & Fenn, 2008). Some predictions were not met, such as Cantonese T5 (23) to Mandarin T3 (214), as Cantonese T5 was identified with poorer accuracy than expected. English listeners identified Cantonese T2 (25) with high accuracy as it may have been mapped to the English question intonation, though the other rising T5 (23) and the falling T4 (21), which was predicted to map onto the English falling intonation, were not identified as consistently. In sum, it appears that there are tone contrasts which are difficult to distinguish across all language groups due to phonetic similarity, as well as contrasts which are difficult to distinguish due to L1 interference, but the mappings between nonnative tones and L1 categories cannot always be correctly predicted (especially for native listeners of a nontonal language).

1.4. Tone training

As outlined in Section 1.2, tone perception is modulated by various factors such as language experience, musicianship, cognitive ability, and hearing and musical disorders. This section evaluates the predictors of successful nonnative tone learning as observed in the tone training literature (for a review, see Antoniou & Chin, 2018). Unlike the studies mentioned in Section 1.2, tone training studies usually involve multiple (at least two) training sessions in order to measure learning improvement from session to session. Some of the more prominent predictors of tone learning success include tone language experience, musicianship, and differences in learning aptitude, which is determined by pretraining pitch sensitivity or overall performance in tone training.

1.4.1 Tone language experience

Findings within the tone training literature have shown that tone language experience generally provides benefits to nonnative tone learning. Native listeners of Thai showed an advantage over native English listeners when learning Cantonese tone words. Tone experience, however, did not bolster performance on the pre- and post-training tone identification tasks; musical experience was a better predictor of success in tone identification. Results suggested that tone identification, a lower level of tone processing, was influenced by both L1 tonal and intonational categories and cue weightings (Cooper & Wang, 2010, 2012). Thai tone training studies have shown similar advantages for learners with native tone experience. Cantonese and Mandarin musicians and nonmusicians outperformed English musicians and nonmusicians in both tone identification and the final session of tone word learning, though Cantonese listeners also outperformed Mandarin listeners in the first training session. This suggests that while tone experience provides advantages for nonnative tone learning, Cantonese listeners have an initial performance boost over Mandarin listeners in tone word identification, possibly due to their more complex (denser) tone inventory (Tong & Tang, 2016). A series of studies examining Thai tone learning by Mandarin and English listeners showed both groups improving significantly in tone identification and discrimination after training, except in the preliminary study, where the English listeners did not show improvement after training. Mandarin listeners consistently outperformed English listeners before and after training, though the English group showed greater improvement in performance after training (Wayland & Guion, 2004; Wayland & Li, 2005, 2008). There is evidence that there is a native tone language advantage for learning a new tone system, and perhaps a tonal complexity advantage, though more research on the effect of complexity on tone training is required.

Differences in tone processing between Mandarin and English listeners have also been observed in neuroscientific studies. English listeners learning Thai tones were more sensitive to early pitch differences and pitch height prior to training. This sensitivity increased following training, shown by an increase in MMN amplitude for high tone deviants and mid standard tones. Conversely, the MMN for high tone deviants decreased for Mandarin listeners after training, indicating that they attended more to later pitch differences and pitch contour than pitch height (Kaan, Wayland, Bao, & Barkley, 2007). In another study, the English group showed more native-like patterns of tone perception following training, but Thai controls still differed from the nonnative Mandarin and English listeners when perceiving late pitch contour differences in Thai, such as high rising and mid level (Kaan, Barkley, Bao, & Wayland, 2008). After Thai tone training, English listeners showed stronger gamma-band activity and alpha-band synchrony than Mandarin listeners as a result of forming new tone categories. Mandarin listeners, who already have native tone categories, require fewer resources to form nonnative tone categories (Kaan, Wayland, & Keil, 2013). It appears that tone training improves nonnative tone processing, possibly with stronger effects for learners with no prior tone experience.

Other studies show that tonal learner groups perform similarly to nontonal groups. Beginner-level Mandarin learners from three different language backgrounds—tonal Hmong, pitch-accented Japanese, and nontonal English—completed either a perception only or perception and production training program. In the tone identification pretest, the Japanese and English groups outperformed the Hmong learners, whose poor performance may have been attributed to their weighting toward the pitch height dimension, as the Hmong tone system contains three level tones. Nevertheless, all groups from both training programs improved significantly in tone identification following training, indicating that tone experience granted neither an advantage nor disadvantage for L2 tone learning (X. Wang, 2013). Similarly, in a Cantonese tone training study, naïve Mandarin and English listeners did not differ significantly in pre- and post-test tone identification or tone word learning. Both groups showed improvement in performance during training and at post-test, and only differed in terms of confusion patterns in the tone identification task (Francis et al., 2008). Once again, these differences were attributed to the listeners' cue weightings and the interaction between nonnative tone and L1 tonal and intonational categories.

Tone training paradigms have consistently been proven to improve tone learning performance for all learner groups, including nontonal listeners with little to no tone experience. Native English listeners who were experienced learners of Thai outperformed the naïve English group on Thai tone discrimination, particularly for open syllables in the shorter ISI condition (500 vs. 1500 ms), though the native Thai listeners outperformed both English groups (Wayland & Guion, 2003). After eight sessions of high-variability Mandarin tone training, beginner-level American English learners of Mandarin showed significant improvement in tone identification from pretest to post-test, improving also in their ability to generalise to new stimuli and new talkers. The learners retained this gain in performance six months following the cessation of training (Y. Wang, Spence, Jongman, & Sereno, 1999). Perceptual training effects transferred over to Mandarin tone production—the pitch contours produced by the English speakers approach nativelike contours following training, but productions are still constrained by perceptual performance (Y. Wang, Jongman, & Sereno, 2003). English learners of Mandarin showed activations in the same locations in the cortex before and after tone training. These areas in the left hemisphere included the supplementary motor cortex, the secondary auditory cortex and both Broca's and Wernicke's areas. Following training, spatial activation in the left superior temporal gyrus increased, and activity also emerged in the right inferior frontal gyrus, signalling the involvement of expanding and recruiting cortical areas in language learning (Y. Wang, Sereno, Jongman, & Hirsch, 2003). In a study measuring the FFR, pitch tracking in naïve English listeners increased following Mandarin tone training and errors decreased, but only for the low dipping tone (214), which the listeners had the least familiarity with compared to the high level (55) and rising (35) tones. This indicates changes in brainstem encoding occur only for stimuli that learners are least familiar with (Song, Skoe, Wong, & Kraus, 2008). Though tone perception is facilitated by prior tone experience, whether native or nonnative, listeners with no tone experience are still capable of learning novel tones through training.

Training paradigms can also impact on successful tone learning. For instance, high variability of training stimuli was most effective when the target phonetic feature (in this case, lexical tone) varied while the irrelevant feature (e.g. voice onset time) did not. Varying both the relevant and irrelevant features had debilitating effects on learning performance (Antoniou & Wong, 2016). Naïve English listeners of Mandarin appear to benefit more from training that utilises orthography with tone marks than orthography without, suggesting that the symbols, while unfamiliar to the listener, facilitate the learning of the novel linguistic feature (Showalter & Hayes-Harb, 2013). English listeners received either perception-only and perception-plus-production training for Mandarin tones, with both groups improving in tone discrimination following training and showing a reduced MMN at post-test. The motor system does not appear to facilitate tone processing, as there was no difference in performance between the learner groups. There is also an asymmetric pattern of tone perception whereby presentation order affects how well tones are discriminated, as certain tone combinations may be heard as native intonational categories (Lu, Wayland, & Kaan, 2015). The training method has further impacts on tone learning when learning aptitude is considered (see Section 1.4.3).

1.4.2 Musical experience

As outlined in earlier sections, musical experience has been proven to facilitate tone perception. English listeners who were amateur musicians, defined as listeners with more than six years of musical training who commenced training before age 10, made up the majority of the successful learners in a Mandarin tone training study. Successful learners consisted of listeners whose tone word learning accuracy surpassed 95% for two consecutive sessions (Wong & Perrachione, 2007). English nonmusicians who received Cantonese tone training performed similarly in tone word learning to musicians who did not receive tone training, with groups outperforming English nonmusicians who also did not receive training. Though there is an initial performance boost granted by long-term musical experience, tone training allows nonmusicians to enhance their perception of lower-level tonal features, as seen in their tone identification performance, in order to improve perception in a higher-level task such as tone word learning (Cooper & Wang, 2011, 2013). Nontone-language-speaking musicians are initially better than their nonmusician counterparts in pitch contour identification only; they did not differ from nonmusicians in pitch contour abstraction and categorisation. Following training, both groups improved significantly in all tasks such that musicians and nonmusicians did not differ in performance. As seen in other tone perception studies, the learners also identified the rising tone with greater accuracy than the falling tone (Wayland, Herrera, & Kaan, 2010). Evidence is mixed regarding the interaction between tone and musical experience on nonnative tone learning. It has been shown that tone experience provides more benefits to learning than musical experience, as tone language listeners with a native tone language can utilise their native categories for learning (Tong & Tang, 2016). Conversely, it has been shown that having either tone or musical experience provides advantages for nonnative tone word learning. Having experience in both a tone language and music, however, does not provide further learning benefits. Also recall that musical experience was shown to better predict successful tone identification than tone experience, indicating that musical experience facilitates the processing of lower-level linguistic information (Cooper & Wang, 2012). While musical experience does provide advantages to tone identification and tone word learning, in some cases, training can also boost nonmusicians' learning performance to that of musicians'.

1.4.3 Learning aptitude

Learning aptitude in tone training studies is often determined by performance in nonlexical pitch pattern identification prior to training, or by overall performance across sessions. English listeners learning Mandarin tones were separated into good or poor learners depending on whether their final training session performance was above or below the median score. Learners who showed more sensitivity to pitch direction, a cue used in Mandarin tone perception, were more successful in learning Mandarin tones, and most of these participants belonged to the good learner group (Chandrasekaran, Sampath, & Wong, 2010). In another study, learning success was determined by performance across consecutive sessions—less successful learners improved by under 5% for four sessions, while more successful learners achieved a score of 95% for two sessions. High sensitivity to pitch contours in the pretraining pitch identification task was shown to facilitate tone identification and word learning, but this only accounted for around half of the variance in attainment level (Wong & Perrachione, 2007). High-variability training can provide better learning outcomes than low-variability training, but only for high-aptitude learners (i.e., listeners who performed above 70% accuracy in pretraining pitch identification); low-aptitude learners who completed high-variability training performed worse than those who completed low-variability training (Perrachione, Lee, Ha, & Wong, 2011). Further findings suggest that low-aptitude learners struggle in high-variability training conditions because they are impaired by cognitive load, whereas high-aptitude learners are not

(Antoniou & Wong, 2015). High-variability training also becomes less effective for learning difficult nonnative categories unless the learners show a high sensitivity to pitch (Sadakata & McQueen, 2014). High-aptitude learners showed more rapid improvement across tone training sessions and outperformed the low-aptitude learners by the end of training, while low-aptitude learners showed greater overall improvement (Ingvalson, Barr, & Wong, 2013). When attempting to maximise tone learning outcomes by introducing high-variability training, it appears that only listeners with higher sensitivity to pitch will show improvements in performance.

1.5. The present study

The present study examined the effect of language experience on nonnative tone perception across five sessions of tone training. The first aim was to compare learners with native tone language experience to those who possessed none. The second aim was to compare learners who are native speakers of different tone languages. Although there are a number of tone perception and tone training studies that have compared learning performance between two or more tone language groups (Alexander & Wang, 2016; Burnham, Kasisopa, et al., 2015; Reid et al., 2015; Tong & Tang, 2016), the majority of studies in this field have compared learning performance between tonal and nontonal language groups (especially when training is involved). Additionally, most tone training studies have presented Mandarin, Cantonese, or Thai tones to learners (Cooper & Wang, 2013; Francis et al., 2008; Kaan et al., 2013; Wayland & Li, 2008; Wong & Perrachione, 2007). Here, we address this gap in the literature by using Hakka as a training language, a regional Chinese dialect not yet used in a tone training program. The three learner groups in this study were native speakers of nontonal Australian English, tonal Mandarin Chinese, and tonal Vietnamese. The below sections provide descriptions of the tonal and intonational inventories of the training language and the learners' languages (see Figure 1 for a visual depiction of the Hakka, Mandarin and Vietnamese tone systems), before presenting the research questions and hypotheses of the thesis.

1.5.1 Hakka Chinese

Meixian Hakka (henceforth Hakka), is the prestige dialect of Hakka Chinese. Most native speakers of the Hakka dialect reside in the southern provinces of Guangdong, Fujian, Jiangxi, Guangxi, Hunan and Sichuan in Mainland China. Other native speakers live in Hong Kong, Taiwan, and parts of South and Southeast Asia such as Malaysia, Singapore, Indonesia and Thailand (Cheung, 2011; Hashimoto, 2010; Saengtummachai, 2003). Hakka is mutually unintelligible with the other varieties of Chinese. Unlike in Mandarin, stop consonant codas are permissible, with two possible checked tones occurring in these contexts. As previously mentioned, these consonants are the unreleased plosives /p̚/, /t̚/, and /k̚/. According to Hashimoto (2010), Hakka's tone system consists of four unchecked "legato" tones (mid level, low level, mid falling and high level) and two checked "staccato" tones (mid falling and high rising). Later studies have described tone 1 (33) as a mid level tone, tone 2 (11) as a low level tone, tone 3 (41) as a mid-high-to-low falling tone, and tone 4 (51) as a high-to-low falling tone, where the level tones are longer in duration than the falling tones. The checked tones, 5 and 6, are shorter than the unchecked tones and consist of a high level tone (55) and a mid-to-low falling tone (41) similar to the unchecked tone 3 (Cheung, 2011; W.-S. Lee & Zee, 2009). In accordance to traditional notations of tonal categories in Middle Chinese, the Hakka tones are also known as the following: *Yin Ping* for tone 1, *Yang Ping* for tone 2, *Shang Sheng* for tone 3, *Qu Sheng* for tone 4, *Yin Ru* for tone 5, and *Yang Ru* for tone 6 (Cheung, 2011), where *Yin* was historically applied to syllables with voiceless initials and *Yang* to syllables with voiced initials (Norman, 1988). Henceforth, the Hakka tones will be referred to as T1, T2, and so on.

As a training language, Hakka's tone system can provide new insights on nonnative tone perception. It has two unchecked level tones that are identified using pitch height, two falling tones that are identified by both pitch height and slope, and two checked tones, one of which is a falling tone (Cheung, 2011). Notably, unchecked T3 and T4 are falling tones that begin at slightly different pitch heights, and this contrast might be difficult for naïve listeners to discriminate due to their similarity. An interesting thing to note is that Hakka has no rising or complex contour tones. Though a mid-to-high rising tone (35) does occur as a tone sandhi form of T1 when it is followed by T2, T3 or T6 (Cheung, 2011), all of the stimuli in this study are monosyllabic and are unaffected by tone sandhi patterns.

1.5.2 Mandarin Chinese

Mandarin is the national language of China and is the largest variety of the Chinese language family. There are four tones, including a neutral tone that occurs in some affixes and non-initial syllables in bisyllabic words (Yip, 2002). The first tone is a high level tone (55), the second is a mid-to-high rising tone (35), the third is a mid-to-low-to-mid dipping tone (214), and the fourth is a high-to-low falling tone (51). The third tone has also been described as a low level and low falling tone with laryngealisation (or creakiness) over the latter half of a syllable (Shih, 1988). While the notation for the first and second tones have been consistent in the literature (C.-C. Chen, 2011; M. Y. Chen, 2000; W.-S. Lee & Zee, 2003; Norman, 1988; Yip, 2002), the third and fourth tones are occasionally presented as 315 (C.-C. Chen, 2011) and 41 (Yip, 2002), respectively. This thesis uses 214 and 51 as the notations for tones 3 and 4, since most studies on Mandarin tone perception follow this convention. Similar to Hakka, the traditional Middle Chinese notations for Mandarin's tonal categories are as follows: *Yin Ping* for T1, *Yang Ping* for T2, *Shang* for T3, and *Qu* for T4 (C.-C. Chen, 2011).

1.5.3 Southern Vietnamese

Vietnamese is a Mon-Khmer language in the Austroasiatic language family. The main varieties of Vietnamese are mostly separated into Northern (Hanoi) and Southern (Ho Chi Minh City) dialects, with the Northern variant being considered the standard dialect of Vietnam. There is also a Central dialect and subdialects in other parts of the country (Brunelle, 2009; Hwa-Froelich, Hodson, & Edwards, 2002). The Vietnamese listeners recruited for this study were all native speakers of the Southern dialect, except for three Central dialect speakers.

The Northern and Southern Vietnamese tone systems differ slightly in density, voice quality and their individual tone contours. The Northern variety has six tones, and some are produced with glottalisation, larygealisation or a breathy voice (Nguyễn & Edmondson, 1997). Contrastingly, the Southern variety has five tones and does not utilise voice quality as a contrastive feature. Both dialects have two checked tones, as the language allows syllables ending with the unreleased plosives /p̚/, /t̚/, and /k̚/. The tones in Southern Vietnamese share the same traditional classifications as the Northern dialect: *ngang* (A1), *huyền* (A2), *sắc* (B1), and *nặng* (B2), while the fifth tone is a tone merger of the *hỏi* (C1) and *ngã* (C2) tones in Northern Vietnamese. The checked tones are variants of the *sắc* (B1) and *nặng* (B2) tones and are labelled D1 and D2 (Brunelle, 2009). In Southern Vietnamese, A1 is a mid-high level tone (44), A2 is a mid-low-to-low falling tone (21), B1 (and D1) is a mid-to-high rising tone (35), B2 (and D2) is a mid-low dipping tone (212), and the fifth tone (referred to as C1 in this study) is a mid-to-low-to-mid dipping tone (214; Kirby, 2010; Vũ, 1982).

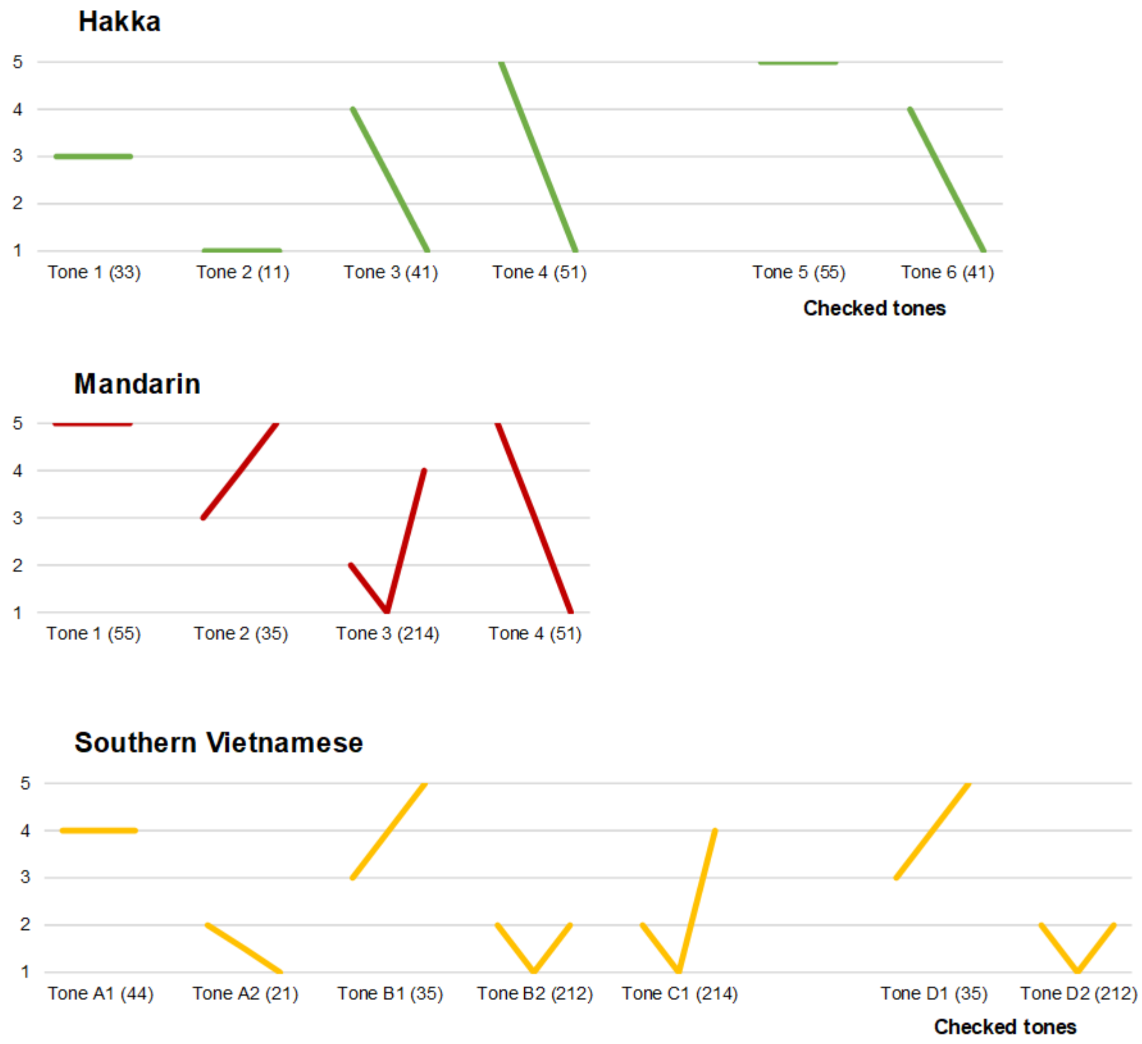

Figure 1. Visualised depictions of the tone systems of Hakka (top), Mandarin (middle), and Southern Vietnamese (bottom), based on previous research. Each tone is described using the Chao (1930) notation system. The pitch targets of a tone, that is, the onset and offset of a tone (and the point where the contour changes direction in a complex tone), are assigned a number corresponding to its pitch height: 1 for the lowest pitch and 5 for the highest. Hakka (adapted from Cheung, 2011) has four tones and two checked tones. Mandarin (adapted from M. Y. Chen,

2000; W.-S. Lee & Zee, 2003; Norman, 1988) has four tones, including a low-dipping tone, but no checked tones. Southern Vietnamese (adapted from Brunelle, 2009; Kirby, 2010; Vũ, 1982) has five tones, two of which are dipping tones, and two checked tones which are allophonic to tones B1 and B2.

1.5.4 (Australian) English

English, which has several varieties such as American, British, and Australian, is a nontone language that utilises pitch at the phrasal sentential level in the form of intonation. There have been several different classifications for the English intonation system in the literature. According to Pierrehumbert and Hirschberg (1990), there are six pitch accents, two phrase accents, and two boundary tones in American English, for a total of 22 contour combinations (Ladd, 2008). Cruttenden (1997) proposes that there are seven nuclear tones in English: low- and high-rise, low- and high-fall, fall-rise, rise-fall, and level (generally mid level). A study examining the perception of Southern British English intonation by Southern British English, Spanish, and Mandarin listeners found that all groups distinguished between terminal falling (HL) and rising (LH) contours in both speech and nonspeech contexts (Grabe, Rosner, García-Albea, & Zhou, 2003). Australian English (AusE), the variety spoken by the English group in this thesis, utilises an intonation distinct from other varieties of English: a high rise tone, known as the high rising terminal (HRT), that occurs at the end of an utterance where a fall is usually expected (Guy, Horvath, Vonwiller, Daisley, & Rogers, 1986). In Australian English, there are nine pitch accents, two phrase accents, and two boundary tones in AusE, as well as five different types of rising intonation (Fletcher, Stirling, Mushin, & Wales, 2002).

Compared to the detailed classifications of intonation patterns above, a smaller number of intonation categories are used to make predictions for nontonal listeners' performance in tone perception and training studies. AusE and French listeners were provided four intonational category options in a Mandarin tone categorisation task: *Flat-pitch* for a level pitch contour; *Question* for a rising contour, as observed in yes-no questions; *Statement* for a falling contour, similar to one at the end of a declarative sentence; and *Exclamation* for a falling contour with a larger pitch excursion (So & Best, 2014). Reid et al. (2015) provided one group of AusE listeners with similar intonation categories, except *Exclamation* was replaced with *Order* and a category labelled *Uncertainty* (i.e., a falling and rising contour) was added. The other AusE listeners were given visual representations of pitch contours as category options. Alternatively, Francis et al. (2008) classifies the intonational categories of American English as follows: a falling intonation contour at the end of a declarative, a high-rise question intonation contour (H*HH%), a yes-no question intonation contour (L*HH%), and a relatively low (L) and relatively high (H) level intonation contour.

1.5.5 Research questions and hypotheses

This thesis is a five-session tone training study that explores the influence of language background on the perception and learning of a nonnative tone system. The following research questions are addressed:

(a) Is nonnative tone learning facilitated by tone language experience?
(b) Does a more "complex" (i.e., denser) native tone system benefit learning?
(c) Can a native language's tonal/intonational features be used to predict which nonnative tones will be successfully learned?

The majority of studies in the literature have shown evidence of tone experience having an overall facilitatory effect on nonnative tone learning rather than no benefit at all. On the basis of this observation, we posit that both Mandarin and Vietnamese groups will outperform the English group in tone identification and tone word learning in the initial and final sessions of the training regimen, as well as in the retention test. If tonal complexity provides further benefits to tone learning, then the Vietnamese group will outperform the Mandarin group, since Vietnamese has five tones and two checked tones compared to Mandarin's four tones. Some studies have shown support for this tonal complexity hypothesis (Tong & Tang, 2016; Zheng et al., 2012). However, if the interaction between the nonnative tones and the learner's native tone/intonation categories predicts learning success, Vietnamese learners may not necessarily outperform Mandarin listeners. In this case, PAM principles can be applied to determine which tones and tone contrasts each learner group will have more difficulty in learning. It should be noted that PAM can only be applied to a certain extent in this study, since this study does not involve tone discrimination *and* categorisation tasks which are required to determine assimilation patterns required to test PAM predictions. However, it is possible to speculate regarding the possible assimilation patterns of learners by observing the similarities between the nonnative tones and the nearest native tone or intonation categories. The weighting of pitch cues, which differs depending on native language, may also predict the nonnative tones which listeners will show more sensitivity to.

Certain Hakka tones and tone contrasts might be difficult for all groups to learn. It has been established that the easiest tone pair to perceive would be a DynamicDynamic (rising–falling) pair, and that tone pairs that include a rising contour would be easier to perceive overall (Burnham, Kasisopa, et al., 2015). Recall that previous studies have shown that rising contour tones are more perceptually salient than falling contour tones (Krishnan et al., 2010b, 2004; Wayland et al., 2015), and the lack of a rising tone in Hakka may pose additional difficulties for nonnative listeners. The only DynamicDynamic tone pair in Hakka is falling–falling, which is hypothesised to be the most difficult tone pair for all groups to distinguish due to their similarity to each other (41 vs 51). Out of the two falling tones, T4 (51) might be easier to identify, as its higher pitch level at onset and larger pitch excursion could be easier for nonnative listeners to perceive. For the T3–T4 contrast, we predict that all learner groups will show confusion patterns similar to the Single Category assimilation type, or Category Goodness if one of the falling tones is identified more predominantly as the other (e.g. T3→T4), while accuracy for the other tone remains high. Two contrasts that all groups may distinguish with higher accuracy are T1–T2 (33 vs. 11) and T2–T4 (11 vs. 51). The tones in the T1–T2 contrast can be somewhat mapped to native tonal categories in Mandarin and Vietnamese, and the English listeners' weightings to the pitch height dimension could facilitate their perception of the two tones. In the T2–T4 tone contrast, both tones begin at different pitch heights (low vs. high), making it easier for the English listeners, who attend more to earlier pitch differences, to distinguish the tones (Kaan et al., 2007). The tone language listeners, who are more sensitive to pitch contour, may also benefit from the dissimilarity between the low tone and the high falling tone. Lastly, higher identification accuracy for the checked tone contrast T5–T6 (55 vs. 41) is predicted, since there are only two possible tones in the checked syllable context.

There are some similarities between Hakka T5 and Mandarin T1 (both 55), Hakka T4 and Mandarin T4 (both 51), and, to a lesser extent, Hakka T2 (11) and Mandarin T3 (214, but recall that T3 is sometimes described as a low or low falling tone). Mandarin listeners have also been shown to map the Cantonese T3 (33) to Mandarin T3 (Francis et al., 2008), and it is possible that this pattern may appear when hearing Hakka T1 (33). These parallels may result in higher tone identification performance for these particular Hakka tones; conversely, the Mandarin group may have trouble distinguishing the Hakka falling T3 and T4 because Mandarin only has one falling tone. Since Hakka T4 is closer to Mandarin T4, the Mandarin group might be able to identify Hakka T4 more consistently than T3. Overall, most Hakka tones appear to map relatively well to Mandarin's tone categories. High accuracy for all tones, but poorer performance in T3 and T4, is expected.

Hakka tones do not map onto Southern Vietnamese tones as closely as Mandarin, but Vietnamese does share similarities with Mandarin's tone system: they both have a rising tone (35) and a dipping tone (214) which is also described as a low level or low falling tone in Mandarin. Vietnamese has a level tone (A1 [44]) that falls somewhere between Hakka T1 (33) and T5 (55), a falling tone (A2 [21]) that begins at a much lower level than Hakka T3 (41), as well as another dipping tone (B2 [212]) which has a lower pitch excursion than C1 (214). For the Vietnamese group, the Hakka level T1, T2 and T5 might be among the easier tones to learn since Hakka T1 and T5 can map onto A1 (44)—and both tones do not occur in the same syllabic contexts, meaning they are less likely to be confused together—and Hakka T2 can map onto both B2 and C1. Similar to Mandarin, there is one falling tone in Vietnamese, but it maps onto neither of the Hakka falling tones. Although the Vietnamese group may also show difficulties distinguishing the falling tones, it would not be because one of the tones is similar to a native category.

Despite the lack of consensus for the number of intonation categories in the English varieties, English listeners at least distinguish between rising and falling intonation contours, as well as level intonation contours. The AusE group may show more consistent identification scores for the level T1 (33), T2 (11) and T5 (55), particularly for T2 and T5, which occur at the lowest and highest pitch levels relative to the other tones. Also recall that nontone language speakers show greater weighting for the pitch height dimension when perceiving tones (Cooper & Wang, 2012; Francis et al., 2008; Guion & Pederson, 2007), suggesting a possible bias towards hearing the Hakka level tones. However, the two Hakka falling tones may cause difficulty for the AusE listeners, given their inexperience with pitch contours in a lexical context. It is possible that Hakka T3 (41) and T4 (51) will map onto the two falling (*Statement* and *Exclamation*) intonation contours, but only if the AusE listeners show enough sensitivity to the differences in pitch contour between the two intonation patterns. Otherwise, it is likely that they will fail to distinguish the two nonnative tones but will show higher identification accuracy for T4, which appears to be the easier tone to perceive.

To test these hypotheses, the learners' performance on tone identification, tone word learning, and generalisation were evaluated. Confusion patterns in tone identification were analysed to determine whether they coincided with the predictions made for each learner group. Results may provide a better understanding of the role tone language experience plays on the learning of a new tone language, and whether tonal complexity adds further advantages, or any advantages at all, to learning. If the successful learning of tones and tone contrasts is better predicted by native tone or intonation categories and language-specific weightings to pitch cues, results may have implications for how we can make more accurate *a priori* judgements for nonnative tone learning.

Chapter 2: Method

2.1. Participants

Participants were separated into three language groups of native speakers: nontonal Australian English, tonal Mandarin Chinese and tonal Vietnamese. There were 22 AusE speakers (*M* age = 24.34, SD = 7.87), 21 Mandarin speakers (*M* age = 26.24, SD = 5.3), and 18 Vietnamese speakers (*M* age = 25.32, SD = 7.78). Data from one participant in the Mandarin group were excluded from analyses (see Results for more information). All participants were free of audiologic or neurological deficits, and all passed an air conduction audiogram set at 25 dB HL at 500, 1000, 2000 and 4000 Hz. Five AusE participants spoke another language at home. These languages included Greek, Arabic and Serbian. Although Greek and Arabic are nontonal, Serbo-Croatian is a pitch-accented language (Yip, 2002). One AusE participant was born in New Zealand and moved to Australia at the age of 10, and also studied Japanese (a pitch-accented language) in high school and university. The Vietnamese speakers were all Southern dialect speakers except for three Central dialect speakers. One of the Southern Vietnamese participants also reported speaking the Nẫu dialect, which is spoken in the South-Central Vietnamese province of Bình Định. Eleven Mandarin speakers spoke another Chinese dialect, but none reported any exposure to Hakka. The Chinese dialects included Cantonese, Shanghainese, Sichuan, Wuxi, Minnan, Henan and Tianjin. Participants who were recruited from Western Sydney University’s Research Participation System were given credits towards their undergraduate psychology course after completing the study. Participants who were recruited from the community were financially reimbursed.

Some participants reported musical experience ranging from 1 to 12 years. There were five participants with musical experience in the AusE group (*M* years of training = 3.9, SD = 4.53), nine participants in the Mandarin group (M = 6.22, SD = 3.38), and six participants in the Vietnamese group (M = 4.17, SD = 2.48).

2.2. Materials

The experiment consisted of three tasks: tone identification, tone word learning, and a generalisation test. The stimuli for the tasks were created from the recordings of three female native speakers of Australian English. Recordings took place in a sound-attenuated booth and were sampled at 44.1 kHz. The talkers produced each token with a level tone and these were recorded using a Shure SM10A cardioid microphone connected to a Roland Duo-Capture EX audio interface. The pitch-synchronous overlap and add function in Praat (Boersma & Weenink, 2018) was used to superimpose the tone onset and offset values taken from three female Hakka speakers (Cheung, 2011), creating new pitch contours for the experimental stimuli. These tone values are shown in Table 1.

Table 1. *Pitch values (Hz) collected from three female native speakers of Hakka at the onset and offset of each tone (Cheung, 2011).*

Tone	Tone numbers	Speaker 1		Speaker 2		Speaker 3	
		Onset	Offset	Onset	Offset	Onset	Offset
T1	33	192	180	238	225	222	215
T2	11	166	131	170	144	201	157
T3	41	210	130	244	140	238	156
T4	51	252	138	290	162	287	163
T5	55	252	265	285	294	276	265
T6	41	217	154	250	152	242	170

The tone values from Speaker 1 were superimposed onto the stimuli recorded for the tone identification task. Speaker 2's tone values were superimposed onto the training task stimuli, and Speaker 3's tone values were superimposed onto the generalisation test stimuli. Two native speakers of Hakka listened to the experimental stimuli and independently judged them to be acceptable exemplars of Hakka tones.

The decision was made to use resynthesised stimuli rather than natural speech tokens for several reasons. Firstly, the estimated pitch values for each of the Hakka tones have already been verified in a prior study involving multiple Hakka native speakers (Cheung, 2011). Second, resynthesised stimuli allowed for the greatest level of experimental control of variables such as duration and intensity; all minimal pairs or tetrads differed only in pitch contour, ensuring that listeners were required to attend to the pitch contours to differentiate the stimuli rather than other cues. Lastly, this method of stimulus creation closely followed the methods used in past tone training studies (Hallé et al., 2004; Wong & Perrachione, 2007).

2.2.1 Tone identification task

For the tone identification task, one talker produced the Hakka monophthongs /a/, /e/, /i/, /o/, and /u/, and the VC syllables /ak̚/, /et̚/, and /ip̚/. Unreleased voiceless plosives were chosen as the syllable codas to better emulate the syllable structure in Hakka. All five Hakka vowels appear within the Australian English vowel space, and the AusE speakers produced the stimuli without difficulty. The four Hakka tones were superimposed onto each of the five monophthongs, while the two checked tones were superimposed onto the three VC syllables. A total of 26 stimuli were created for the tone identification task. The tokens /a/ and /ak̚/ were presented as practice stimuli, and the remaining tokens were used as test stimuli. Stimuli containing checked tones appeared in the VC syllable context rather than CVC as the goal of this task was to identify the pitch contour rather than to attend to the segments—a VC syllable structure provides the simplest context available for a Hakka checked tone to appear. For this reason, it has been argued that such a task is nonlexical as the focus is on pitch differences (Chandrasekaran et al., 2010). Since the checked tone stimuli in the training and generalisation tasks were CVC syllables, the tone identification stimuli were also made distinct to determine whether there were any gains for learners when exposed to a novel phonetic context in a novel phonetic task.

As is shown in Figure 2, arrows were created as a visual aid for the tone identification task. Each arrow represented one of the Hakka tones as they would appear on Chao’s (1930) five-level tone space, with tone height being depicted along the vertical axis and direction along the horizontal axis.

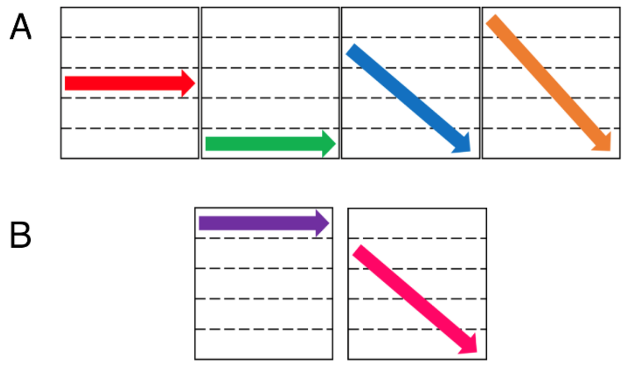

Figure 2. Arrows depicting the height and direction of each Hakka tone. **A.** Unchecked tones are shown in top four panels: Tone 1 (33) mid level tone (top left panel), Tone 2 (11) low level tone (top centre left panel), Tone 3 (41) mid-high-to-low falling tone (top centre right panel), Tone 4 (51) high-to-low falling tone (top right panel). **B.** Checked tones appear in the two bottom panels: Tone 5 (55) high level tone (bottom left panel), Tone 6 (41) mid-to-low falling tone (bottom right panel).

2.2.2 Tone training task and generalisation test

The remaining talkers produced the CVC nonsense words [fon], [leŋ], [nun], [wap̚], and [mip̚] for the tone training task and the generalisation test. Four unchecked tone stimuli were created from [fon], [leŋ], and [nun], while two checked tone stimuli were created for [wap̚] and [mip̚]. This resulted in a total of 16 training stimuli for each task. All nonsense words fell within the phonotactic constraints of English, Mandarin and Vietnamese, and were also nonsense words in all three languages. The only exceptions were [wap̚] and [mip̚], which are not permissible syllables in Mandarin but were required to create the checked tone stimuli.

The auditory stimuli were paired with monochromatic pictures of high-frequency words. These pictures were taken from a set used in past tone training studies (Antoniou & Wong, 2016;

Chandrasekaran et al., 2010; Wong & Perrachione, 2007). The sixteen sound-to-image combinations are shown in Table 2.

Table 2. *Sound-to-image combinations for tone training stimuli*

Tone	Word				
	fon	leŋ	nun	wap̚	mip̚
T1 (33)	hat	piano	scissors		
T2 (11)	bed	cookie	knife		
T3 (41)	cat	plane	ball		
T4 (51)	apple	boat	pencil		
T5 (55)				baby	glass
T6 (41)				dog	brush

2.3. Procedure

The training protocol consisted of five training sessions held on separate days and an optional retention test held at least three weeks after session five. In the first session, participants completed demographic questionnaires, including the Language Experience and Proficiency Questionnaire (LEAP-Q; (Marian, Blumenfeld, & Kaushanskaya, 2007). They then completed the tone identification pretest and their first session of tone training. Session 1 lasted approximately 50 minutes. Sessions 2 to 4 were 20 minutes each and consisted of only the tone training task. In session 5, participants completed their final session of training, followed by a generalisation test and a tone identification post-test. The post-test was identical to the tone identification pretest and was used to measure improvement in tone identification accuracy following training. Session 5 lasted 40 minutes. A total of 86% participants (64% from the AusE group, 100% from the Mandarin group and 94% from the Vietnamese group) returned on a separate sixth day to complete the 10-minute retention test. The retention test comprised of a third administration of the tone identification task whose purpose was to determine whether tone identification performance was retained three weeks after the cessation of training. The sessions were conducted in a quiet testing space. Tasks were presented on a laptop PC running E-Prime 3.0. Audio stimuli were presented using Sennheiser HD280 Pro headphones with stimulus output calibrated to 72 dB SPL.

2.3.1 Tone identification task

The tone identification task was completed in sessions 1 and 5 as a pre- and post-test, and in session 6 as a retention test. The task was approximately 10 minutes long. Prior to commencing the task, participants were first introduced to a visual representation of the Hakka tones (see Figure 2). Participants from the AusE group were provided with further explanation of lexical tone to ensure they understood the task's instructions. The task consisted of a familiarisation, practice and test phase. The familiarisation phase involved exposure to three repetitions of the sound /a/ in four unchecked tones and /ak̚/ in two checked tones. Each sound was paired with an arrow which participants were instructed to memorise. The trials were presented with an interstimulus interval (ISI) of 3 seconds. During the practice phase, participants listened to the same sounds and were instructed to identify the correct tone from the two arrows presented on the screen. Participants pressed '1' on the keyboard for the option on the left of the screen and '0' for the option on the right. Feedback was provided after each trial; the correct response was shown on the screen when participants answered correctly.

The test phase followed a similar procedure with a different set of auditory stimuli produced by the same talker. Feedback was not provided this time and participants were given 3 seconds to respond to each trial. Participants were exposed to every possible tone contrast four times. There were six combinations per unchecked tone and two per checked tone, for a total of 104 test trials. In both practice and test phases, the positions of the response options were counterbalanced. In all phases of the task, the checked and unchecked stimuli were separated into blocks and the order of presentation (checked then unchecked, or vice-versa) was randomised. Within each block, the trials were also presented randomly. The different phases of the tone identification task are depicted in Figure 3 below.

Tone identification performance was calculated by the mean accuracy of responses per session, which also included trials which participants failed to respond to. To analyse the pattern of responses made for each tone, confusion matrices were created. Further analyses were made on response patterns for separate tone contrasts to determine whether specific tone contrasts posed more difficulties for learners.

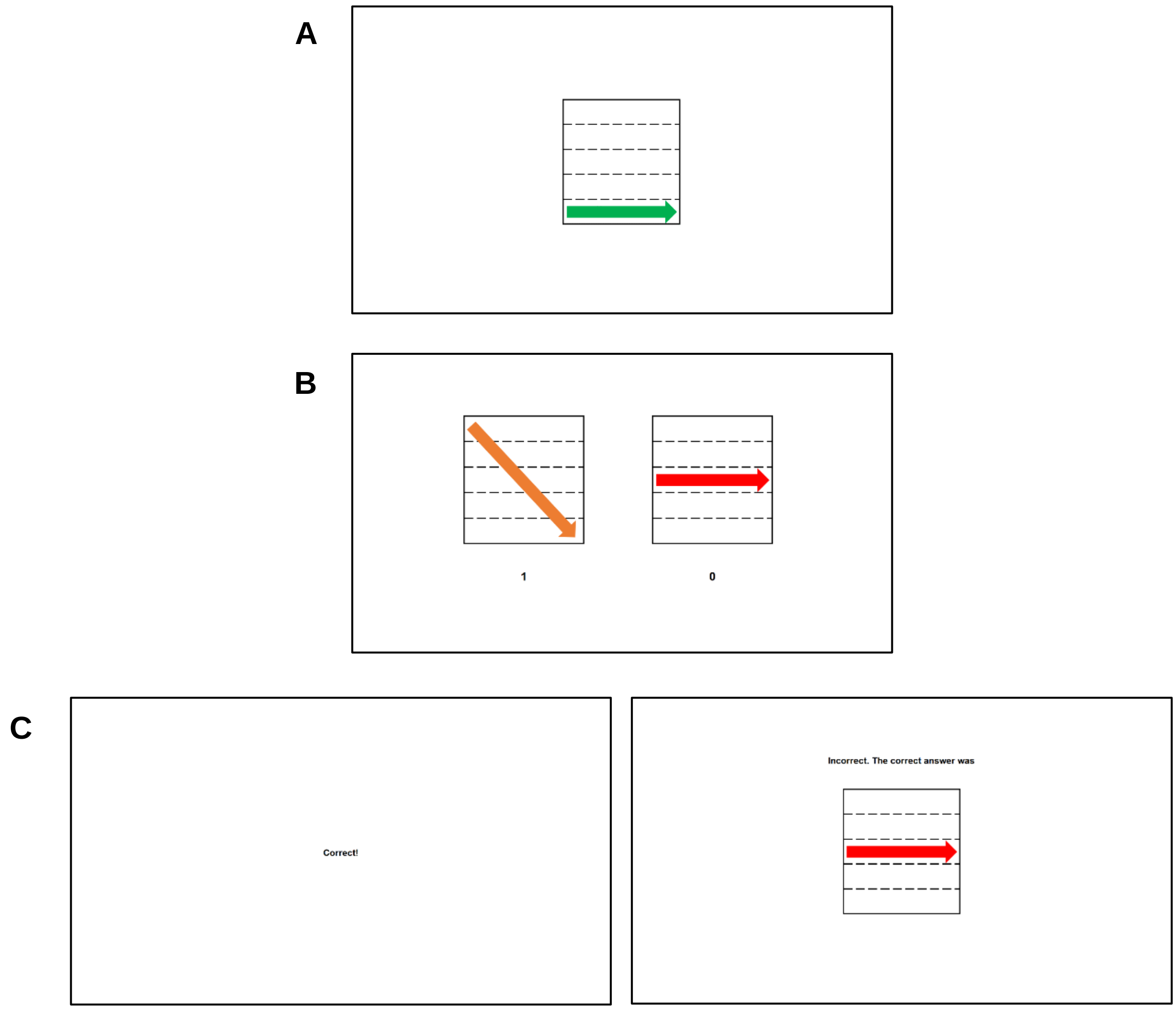

Figure 3. Visual layout of the familiarisation (**A**), practice (**B** and **C**) and test (**B**) phases of the tone identification task. Note that the practice and test phases are identical except that feedback (**C**) was not provided during test.

2.3.2 Tone training task

Participants completed tone word learning tasks in sessions 1 to 5. The word learning task consisted of a 15-minute training phase and a 5-minute test phase. During the training phase, participants completed a familiarisation and practice block for all possible minimal pair combinations. These blocks were similar to those in the tone identification task. There were six minimal pair combinations for [fon], [leŋ] and [nun], and two combinations for [wap̚] and [mip̚]. In total, there were 20 familiarisation–practice blocks. These blocks were presented at random, and unlike the tone identification task, checked and unchecked stimuli were not separated into blocks. Within a single familiarisation–practice block, participants were exposed to four repetitions of a minimal pair, e.g., [fon1] and [fon2], presented in random order. Each trial was presented with an ISI of 3 seconds and consisted of an auditory stimulus and its corresponding image, as outlined in Table 2. In the practice phase, participants listened to a sound and identified the correct image out of two options on the screen. Four trials were presented for each sound, and feedback was provided for each response. As seen in Figure 4, the training phase is similar to that of the familiarisation and practice phases of the tone identification task.

During the test phase, participants heard an auditory stimulus and were required to choose the correct image between the 16 options on the screen. Underneath each image was a letter indicating the key to press on the keyboard. The 16 words were repeated four times, for a total of 96 test trials. Although each trial could only be heard once, participants were not required to respond within a certain time limit. The next trial was presented only once a response was made.

Tone word learning performance was measured by the mean response accuracy in each training session. Alongside overall performance in the task, types of errors were examined to determine whether the errors made in training were tone-only, segment-only or both tone and segment errors. Confusion matrices and tone contrast analyses were also made for training results.

2.3.3 Generalisation test

The generalisation test was held after the training task in session 5. This 5-minute task tested participants on how well they could adapt to a novel talker's speech. The test consisted only of the test phase of the tone training task (differing only in talker) and lasted 96 trials.

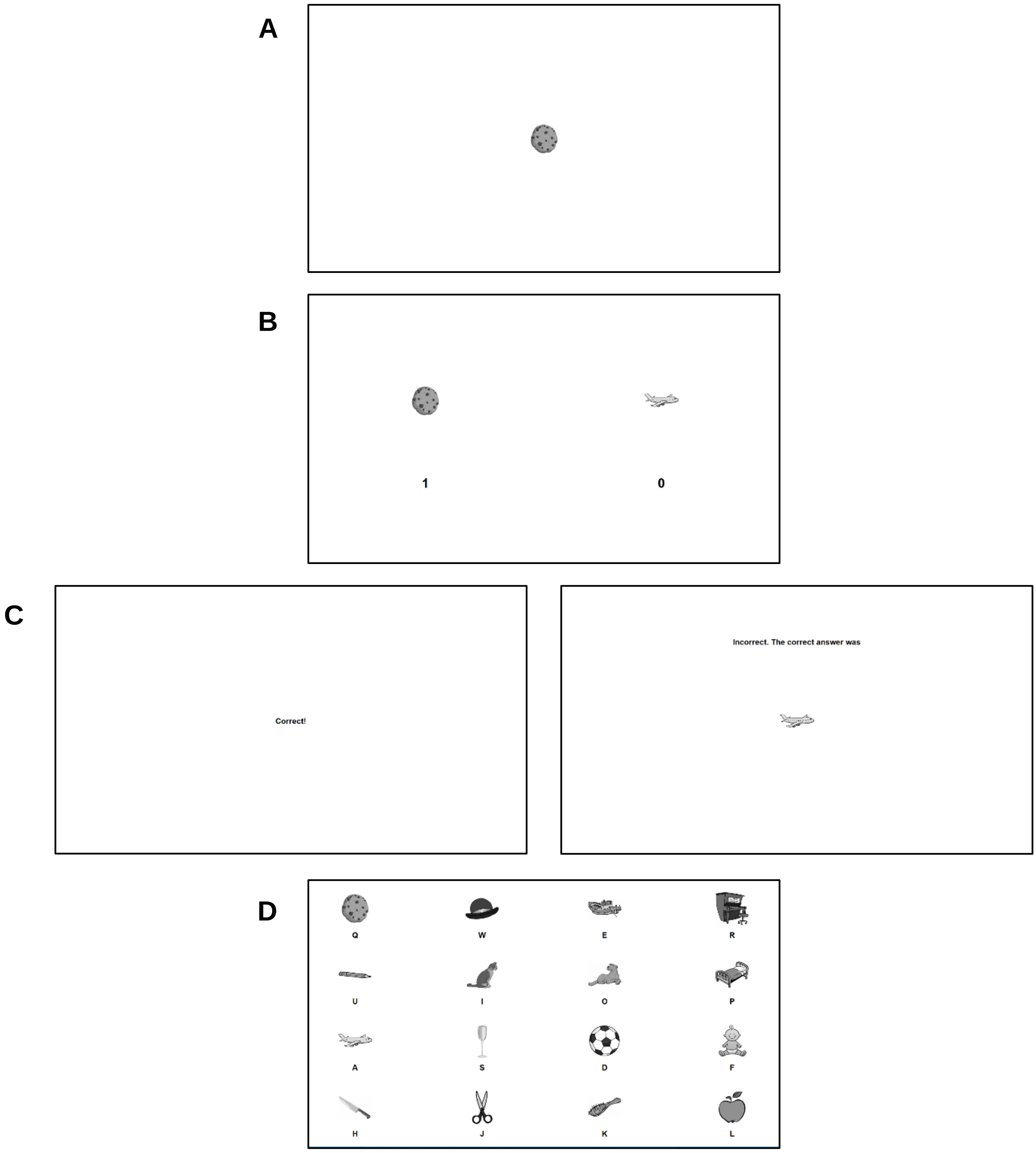

Figure 4. Visual layout of the training (**A**, **B**, **C**) and test (**D**) phases of the tone word learning task. The generalisation test was identical in design to the tone word learning test.

Chapter 3: Results

Adjustments to the data were made prior to conducting statistical analyses. Scores from one participant in the Mandarin group were excluded because the participant failed to properly follow instructions. Outliers in the tone identification tasks were adjusted in accordance to guidelines set by Tabachnick and Fidell (2007) to one unit beyond the next most extreme score. In the tone identification pretest, data from one Mandarin participant and one English participant were adjusted, and in the post-test, one English participant's score was adjusted.

3.1. Tone identification pretest, post-test and retention test performance

3.1.1 Tone identification accuracy

We compared the groups' performance in tone identification as measured by the mean percentage of correctly identified tones. Figure 5 depicts tone identification results at pretest, post-test, and retention test. It appears that all groups showed improvement from pre- to post-test and maintained their training gains in the retention test held three weeks after cessation of training.

To assess the learners' tone identification performance in the pre- and post-tests, results were submitted to a 3 × (2) ANOVA[1] with language (AusE vs. Mandarin vs. Vietnamese) as the between-subjects factor and test (pre- vs. post-test) as the within-subjects factor. Due to the low percentage of participation in the AusE group for the retention test (64% compared to 100%

[1] Since there were several participants in this study who reported musical experience, we also ran ANCOVAs on all tasks, with years of musical training as a covariate. Main effects of language and test/session that were observed in the ANOVAs were still significant when musical experience was included as a covariate, showing that musical experience in this study had no significant effect on tone identification and tone learning performance. However, when comparing post-test and retention test performance between the Mandarin and Vietnamese groups, the main effect of language was not significant when musical experience was a covariate (see footnote 2).

and 94% in the Mandarin and Vietnamese groups), retention test performance could not be compared alongside the pre- and post-tests. A significant main effect was found for test, $F(1, 57) = 13.647$, $p < .001$, $\eta_p^2 = .193$, indicating that tone identification improved following training ($M_{\text{pre}} = 69.7\%$ vs. $M_{\text{post}} = 75.6\%$). There was also a significant main effect of language, $F(2, 57) = 19.179$, $p < .001$, $\eta_p^2 = .402$). Post-hoc Sidak analyses revealed that both tonal groups outperformed the nontonal AusE group (Mandarin, $M_{\text{Diff}} = 21.8$, 95% CI [12.1, 30.5], $p < .001$; Vietnamese, $M_{\text{Diff}} = 10.4$, 95% CI [1.5, 19.3], $p = .017$). The Mandarin group also outperformed the Vietnamese group ($M_{\text{Diff}} = 11.4$, 95% CI [2.3, 20.5], $p = .009$). There was no significant interaction between test and language, $F(2, 57) = .463$, $p = .632$, $\eta_p^2 = .016$).

An insufficient number of AusE participants completed the retention test ($n = 14$), and therefore an analysis was conducted on the retention test data using only scores from the Mandarin and Vietnamese groups. A 2 × (2) ANOVA with language (Mandarin vs. Vietnamese) as the between-subjects factor and test (post-test vs. retention test) as the within-subjects factor revealed a main effect of language[2], $F(1, 35) = 4.360$, $p = .044$, $\eta_p^2 = .111$, with the Mandarin group outperforming the Vietnamese group ($M_{\text{Diff}} = 7.5$, 95% CI [.2, 14.7], $p = .044$). There was no significant difference in test, $F(1, 35) = 1.410$, $p = .243$, $\eta_p^2 = .039$, showing that there was no decline in tone identification performance three weeks after the cessation of training.

[2] This effect was not significant when music was included as a covariate, $F(1, 34) = 3.478$, $p = .071$, $\eta_p^2 = .093$. Recall in Section 2.1 that there were nine participants with musical experience in the Mandarin group (*M* years of training = 6.22, SD = 3.38), and six participants in the Vietnamese group (*M* years of training = 4.17, SD = 2.48).

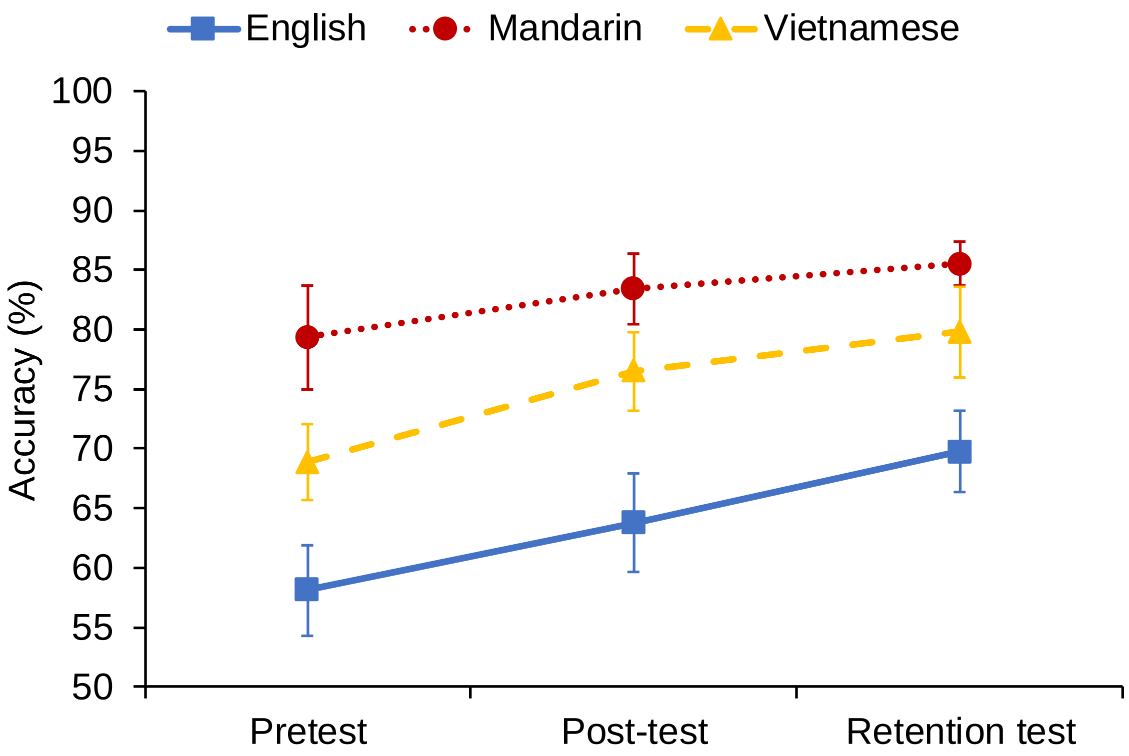

Figure 5. Tonal (Mandarin, Vietnamese) and nontonal (AusE) learners' tone identification accuracy (%) prior to the commencement of training (pretest), immediately following the cessation of training (post-test), and three weeks after training (retention test). Error bars represent standard error of the mean.

3.1.2 Confusion matrices

Confusion matrices were created to examine the learners' tone identification patterns, including which tones learners selected instead of target tones. Recall that for the tone identification task, checked and unchecked tone stimuli were presented in separate blocks, so it was not possible for learners to confuse the checked tones with the unchecked tones. As early as the pretest, all groups had correctly identified the target tone at least 50% of the time. Confusion patterns for each group are shown at pretest (Table 3), post-test (Table 5) and retention test (Table 7).

Additionally, in Tables 4, 6 and 8, we present a breakdown at these same three time points of tone identification accuracy scores for each target tone by each possible visual distractor tone (e.g., when participants heard Tone 1, we present tone identification accuracy when the visual distractor was Tone 2 vs. Tone 3 vs. Tone 4). These tables provide insight into which tone pairs learners found more difficult to differentiate.

Table 3. *Mean tone identification accuracy (%) by AusE, Mandarin, and Vietnamese listeners at pretest. The speaker's intended tone category label is bolded.*

Group	Stimulus tone	Selected tone label					
		T1 (33)	T2 (11)	T3 (41)	T4 (51)	T5 (55)	T6 (41)
AusE	T1 (33)	**67**	10	12	12		
	T2 (11)	10	**70**	9	11		
	T3 (41)	18	14	**55**	12		
(Missed 8.4% of total trials)	T4 (51)	14	8	17	**61**		
	T5 (55)					**74**	26
	T6 (41)					39	**61**
Mandarin	T1 (33)	**90**	5	3	2		
	T2 (11)	3	**83**	8	6		
	T3 (41)	6	8	**80**	5		
(Missed 5.6% of total trials)	T4 (51)	4	3	10	**83**		
	T5 (55)					**94**	6
	T6 (41)					14	**86**
Vietnamese	T1 (33)	**79**	7	7	6		
	T2 (11)	7	**74**	10	8		
	T3 (41)	13	13	**63**	11		
(Missed 4.3% of total trials)	T4 (51)	10	7	14	**69**		
	T5 (55)					**83**	17
	T6 (41)					19	**81**

Although all groups of learners selected the speaker’s intended tone label more than 50% of the time at pretest (the lowest was 55% for AusE listeners’ identification of T3), identification scores varied considerably across tones and groups. Across all groups, the level tones (particularly T1 and T5) were identified with high accuracy, while the falling T3 was seemingly identified less consistently. The AusE group had the lowest overall tone identification accuracy, and their highest identification scores were for the level tones 1, 2 and 5. This coincides with past observations where English native speakers showed high identification accuracy for low, mid, and high level tone contours (Francis et al., 2008). The Mandarin group’s tone identification performance for all tones was high (at least 80% accuracy), even for T3 and T4, which they were predicted to have trouble distinguishing since Mandarin contains only a single falling tone. Similar to the AusE group, the Mandarin group’s accuracy was highest for the level tones 1 and 5. The Vietnamese group did not seem to show an advantage in checked tone identification accuracy despite having checked tones in their native tone system—though they identified the checked tones consistently (T5 = 83%, T6 = 81%), the Mandarin group’s scores were still higher (T5 = 94%, T6 = 86%). Furthermore, it was expected that accuracy in the T5–T6 contrast would be higher amongst all groups, given that there was only one visual distractor for this contrast compared to three for the unchecked tones. Regardless, the AusE group’s accuracy for T6 was relatively low (61%).

To make further observations on the learners’ confusion patterns for every possible tone contrast, we also examined identification accuracy for tones when the target was paired with different visual distractor tones. The visual distractor tone was the alternate, incorrect response option during the tone identification test phase. Table 4 examines the learners’ identification accuracy for every possible tone contrast in the task:

Table 4. *Mean tone identification accuracy (%) at pretest presented separately for each visual distractor that appeared as a response option. The checked tone contrast T5–T6 is omitted since there was only one possible alternate response option.*

Group	Stimulus tone	Visual distractor tone			
		T1 (33)	T2 (11)	T3 (41)	T4 (51)
AusE	T1 (33)		71	65	64
	T2 (11)	69		74	67
	T3 (41)	46	58		63
	T4 (51)	58	76	48	
Mandarin	T1 (33)		85	93	94
	T2 (11)	90		76	82
	T3 (41)	82	75		83
	T4 (51)	89	90	70	
Vietnamese	T1 (33)		78	79	81
	T2 (11)	78		69	75
	T3 (41)	61	59		68
	T4 (51)	71	78	57	

Tone identification accuracy for some stimulus tones increased or decreased when paired with particular visual distractors, showing that there were certain contrasts that the learners found easier or more difficult to distinguish. All learner groups showed high identification accuracy for T4 in the T2–T4 contrast, which may be due to the large difference in pitch height at the onset of the tones. In the T3–T4 contrast, there was a tendency for T3 to be identified more accurately by all groups, indicating that T4 may be more easily confusable as T3. The AusE group also showed high identification accuracy for T1 in the T1–T2 contrast and T2 in the T2–T3 contrast, but poor accuracy for T3 in the T1–T3 contrast. This suggests that the falling T3 was perceived by the AusE group as a mid level tone. Both the Mandarin and Vietnamese groups identified T1 with high accuracy regardless of visual distractor and showed highly consistent identification for T2 when T1 was presented as the visual distractor. The tonal groups' lowest identification scores were for the T2–T3 and T3–T4 contrasts. Interestingly, the Mandarin group was able to identify T3 (41) with higher accuracy than T4 in the T3–T4 contrast (83% vs. 70%), even though Hakka's falling T4 (51) would map onto Mandarin's falling T4 (51) more closely. For both tonal groups, T3 was more confusable in the T2–T3 contrast than in the T3–T4 contrast. Despite the very small differences between T3 (41) and T4 (51) in Chao notation, these identification scores may indicate that the falling T3 is sometimes confused as a low level tone, whereas T4 is heard more as a falling tone, except when T3 is the visual distractor. One would assume that the tone with the greater pitch excursion (T4) would be identified more often as the falling tone, yet this does not appear to be the case at pretest.

Table 5. *Mean tone identification accuracy (%) by AusE, Mandarin, and Vietnamese listeners at post-test.*

Group	Stimulus tone	Selected tone label					
		T1 (33)	T2 (11)	T3 (41)	T4 (51)	T5 (55)	T6 (41)
AusE	T1 (33)	**69**	7	11	12		
	T2 (11)	6	**78**	7	8		
	T3 (41)	16	12	**64**	8		
(Missed 6.6% of total trials)	T4 (51)	15	8	14	**63**		
	T5 (55)					**77**	23
	T6 (41)					29	**71**
Mandarin	T1 (33)	**94**	3	1	1		
	T2 (11)	2	**88**	7	3		
	T3 (41)	4	8	**79**	10		
(Missed 3.5% of total trials)	T4 (51)	3	3	10	**84**		
	T5 (55)					**95**	5
	T6 (41)					12	**88**
Vietnamese	T1 (33)	**83**	4	6	6		
	T2 (11)	6	**79**	9	6		
	T3 (41)	11	11	**70**	8		
(Missed 1.7% of total trials)	T4 (51)	9	5	8	**78**		
	T5 (55)					**82**	18
	T6 (41)					20	**80**

Identification accuracy for tones showed general improvement across groups following training. The AusE group showed great improvement in identifying T2, T3 and T6. Although the Mandarin group's overall performance was high, they showed the most notable improvement in T1 and T2, while the Vietnamese group improved in identifying T3 and T4. Most patterns observed at pretest also remained at post-test. The checked tones T5 and T6 were still identified with high accuracy across all groups. Accuracy for the level tones T1 and T2 remained high for all groups, but the improvement in accuracy for T1 in the AusE group was minimal. All groups also continued to struggle with identifying T3 compared to the other tones; however, the AusE group also showed poor identification accuracy for T4.

Table 6. *Mean tone identification accuracy (%) at post-test presented separately for each visual distractor that appeared as a response option.*

Group	Stimulus tone	Visual distractor tone			
		T1 (33)	T2 (11)	T3 (41)	T4 (51)
AusE	T1 (33)		79	66	63
	T2 (11)	81		79	75
	T3 (41)	52	64		77
	T4 (51)	56	76	58	
Mandarin	T1 (33)		91	97	96
	T2 (11)	93		80	90
	T3 (41)	98	77		71
	T4 (51)	91	92	69	
Vietnamese	T1 (33)		87	82	81
	T2 (11)	82		72	82
	T3 (41)	68	66		76
	T4 (51)	72	85	77	

The T1–T2 and T2–T4 contrasts remained the most easily distinguishable contrasts across all groups. In the T1–T2 contrast, it appears that the pitch heights of both level tones are distinct enough to be discriminable, and as mentioned earlier, listeners may be able to attend to the difference in pitch height between T2 and T4 at tone onset in order to distinguish them. While the Mandarin and Vietnamese groups identified T1 with high accuracy regardless of distractor, the AusE group had difficulty correctly identifying the falling T3 and T4 when the visual distractor was the mid level T1, and their accuracy for T1 also decreased when T3 or T4 were distractors. In all groups, identification accuracy for T3 was not as high as T4 when the low level T2 was the distractor. Similar to scores at pretest, this pattern suggests that the falling T3 is more easily confusable as a low level tone. Although the AusE group’s identification for T4 in the T3–T4 contrast remained relatively inconsistent, their accuracy for T3 in this contrast was high. In the Mandarin group, the T1–T3 contrast was identified with high accuracy, though both the AusE and Vietnamese groups showed poor identification accuracy for T3 in this contrast. As observed in the pretest, the Mandarin group’s accuracy for the T2–T3 and T3–T4 contrasts remained poorer than the other tone contrasts. Training did not facilitate their tone identification accuracy for the T3–T4 contrast, with accuracy for T3 seemingly decreasing at post-test. It seems that following training, the Mandarin group’s confusion patterns have begun to match the predictions made in the hypothesis: the single falling tone in Mandarin would cause the group to struggle to distinguish the two falling tones in Hakka. The Vietnamese group had difficulty identifying T3 when the visual distractors were T1 or T2, indicating that the falling T3 was more confusable as a mid or low level tone. Unlike the other groups, the Vietnamese group’s identification accuracy for the T3–T4 contrast was relatively high for both tones, showing that they were able to distinguish the two falling tones more consistently.

Table 7. *Mean tone identification accuracy (%) by AusE, Mandarin, and Vietnamese listeners at retention test.*

Group	Stimulus tone	Selected tone label					
		T1 (33)	T2 (11)	T3 (41)	T4 (51)	T5 (55)	T6 (41)
AusE	T1 (33)	**72**	6	12	10		
	T2 (11)	5	**80**	8	8		
	T3 (41)	16	12	**62**	10		
(Missed 0.8% of total trials)	T4 (51)	12	7	13	**67**		
	T5 (55)					**91**	9
	T6 (41)					28	**72**
Mandarin	T1 (33)	**95**	2	2	2		
	T2 (11)	4	**86**	7	4		
	T3 (41)	3	6	**80**	11		
(Missed 0.5% of total trials)	T4 (51)	3	3	11	**83**		
	T5 (55)					**98**	2
	T6 (41)					17	**83**
Vietnamese	T1 (33)	**86**	4	5	5		
	T2 (11)	5	**83**	8	3		
	T3 (41)	10	11	**72**	7		
(Missed 1.2% of total trials)	T4 (51)	7	5	7	**81**		
	T5 (55)					**84**	16
	T6 (41)					17	**83**

The retention test assessed tone identification performance three weeks after completion of the post-test. The scores suggest that there was no loss in performance following the cessation of training. Recall that all of the participants in the Mandarin and Vietnamese groups (except for one Vietnamese speaker) returned for the retention test, but only 14 AusE participants (64%) completed this test. Therefore, the AusE confusion matrix may not fully represent the identification patterns for the entire learner group. Parallel to the results from the pre- and post-tests, all groups showed high identification performance for the level tones T1 (albeit to a lesser extent for the AusE group), T2 and T5, and were able to distinguish the checked tones consistently. The Mandarin group identified all tones with high accuracy (≥ 80%), but the group's accuracy was lowest for the falling tones T3, T4 and T6. The AusE and Vietnamese groups once again showed poorer identification for T3, with the AusE group also showing poor identification accuracy for T4. It appears that the level tones were the easiest to identify for all learner groups, while the falling tones, particularly T3 (41), were identified with less accuracy.

Table 8. *Mean tone identification accuracy (%) at retention test presented separately for each visual distractor that appeared as a response option.*

Group	Stimulus tone	Visual distractor tone			
		T1 (33)	T2 (11)	T3 (41)	T4 (51)
AusE	T1 (33)		83	64	69
	T2 (11)	86		77	75
	T3 (41)	53	64		71
	T4 (51)	64	78	60	
Mandarin	T1 (33)		95	95	94
	T2 (11)	89		80	89
	T3 (41)	92	82		66
	T4 (51)	90	92	68	
Vietnamese	T1 (33)		88	85	85
	T2 (11)	84		75	90
	T3 (41)	71	67		79
	T4 (51)	80	85	78	

Confusion patterns in the retention test were similar to those observed at post-test: all groups distinguished the T1–T2 and T2–T4 contrasts consistently. The Mandarin and Vietnamese groups showed high identification accuracy for both T1 and T2, regardless of the visual distractor they were paired with. While the AusE group also showed high identification accuracy for T2 in all contexts, their identification accuracy for T1 was poor when the visual distractors were falling tones. Likewise, when either T3 or T4 were presented with T1 as the distractor, identification accuracy also declined. Both the AusE and Mandarin groups continued to show difficulty in distinguishing the T3–T4 contrast, while the Vietnamese group found the T2–T3 contrast most difficult to differentiate. The falling T3 was more easily confused as either the mid level T1 by the AusE and Vietnamese groups, or the low level T2 by all learner groups. This may be due to the lower starting pitch height and pitch excursion for T3 (41) compared to T4 (51), with T4 being more accurately identified than T3 when the visual distractor was a level tone. In the case of the AusE and Vietnamese groups, T3 could be perceived as a level tone that is not as distinct to T1 (33) or T2 (11) as both tones are to each other. This might also explain why their accuracy for T3 was higher when the distractor was T4. The opposite was observed for the Mandarin group. They could be perceiving T3 as a falling tone, since their identification accuracy was higher when level tones were the visual distractor. As a result, when another falling tone was the distractor, the Mandarin group's accuracy decreased substantially. Interestingly, from pretest to retention test, the Mandarin group's accuracy in the T3–T4 contrast worsened only for T4, while accuracy for T3 remained unchanged. It is possible that, as naïve listeners in the pretest, the Mandarin listeners were able to detect the subtle differences in pitch height and excursion between the two falling tones, but prolonged exposure to the tones caused the group to conflate them to a single falling category.

3.2. Tone training and generalisation test performance

3.2.1 Tone word learning accuracy

Tone word learning performance was calculated by the proportion of correct responses in the test phase of the training tasks. The learners' performance across the five tone training sessions (and the generalisation test) are depicted in Figure 6. Performance in the five training sessions was compared via a 3 × (5) ANOVA with the between-subjects factor of language (AusE vs. Mandarin vs. Vietnamese) and the within-subjects factor of session (1 to 5). There was a main effect of session with a Huynh-Feldt adjustment to the degrees of freedom, $F(2.544, 145.002) = 173.042$, $p < .001$, $\eta_p^2 = .752$), indicating that performance improved across training sessions. A series of post-hoc *t*-tests revealed that performance significantly improved across sessions 1 through 4 ($p < .001$) but was only marginally significant between sessions 4 and 5 ($p = .089$). There was also a significant main effect of language, $F(2, 57) = 9.804$, $p < .001$, $\eta_p^2 = .256$). Sidak post-hoc analyses showed that Mandarin learners outperformed the AusE learners ($M_{\mathrm{Diff}} = 22.3$, 95% CI [9.9, 34.7], $p < .001$), but there was no significant difference between the Vietnamese and AusE groups ($M_{\mathrm{Diff}} = 10.6$, 95% CI [-2.1, 23.3], $p = .130$) and only a marginally significant difference between the Mandarin and Vietnamese groups ($M_{\mathrm{Diff}} = 11.7$, 95% CI [-1.3, 24.7], $p = .09$). There was no significant interaction between session and language, $F(5.088, 145.002) = 1.029$, $p = .403$, $\eta_p^2 = .035$).

We next compared the groups' performance on the generalisation test. Recall that generalisation involved the same procedure as the test phase of the training task but the generalisation stimuli were produced by a previously unencountered talker. A one-way ANOVA was conducted to examine learners' accuracy in the generalisation test, with language as the between-subjects factor. There was a main effect of language, $F(2, 57) = 12.601$, $p < .001$, $\eta_p^2 = .307$). Sidak post-hoc analyses showed that the Mandarin group outperformed the AusE group ($M_{\mathrm{Diff}} = 23.3$, 95%

CI [11.6, 35.0], $p < .001$) and the Vietnamese group also outperformed the AusE group (M_{Diff} = 15.8, 95% CI [3.7, 27.8], $p = .006$). Mandarin and Vietnamese learners did not differ (M_{Diff} = 7.6, 95% CI [-4.7, 19.9], $p = .353$).

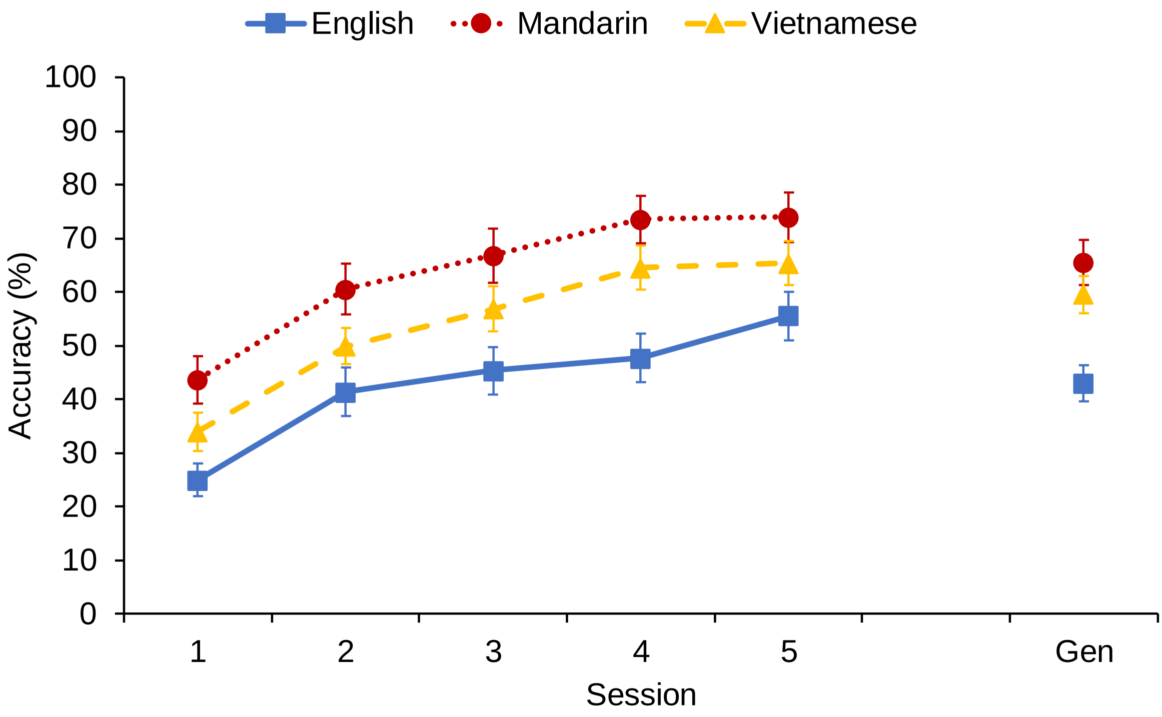

Figure 6. Tonal (Mandarin, Vietnamese) and nontonal (AusE) learners' word identification accuracy over five sessions of training and generalisation test (Gen). Error bars depict SEM.

There were 16 response options in the training task and generalisation test. Three possible errors can be made by learners (Wong & Perrachione, 2007): segment-only errors, where the tone is identified correctly but not the phonetic segment (e.g., identifying the target word /fon1/ as /leŋ1/); tone-only errors, where the phonetic segment is identified correctly but not the tone (e.g., identifying the target word /fon1/ as /fon2/); and errors where both segment and tone were identified incorrectly (e.g., identifying the target word /fon1/ as /leŋ2/). To assess the difficulties encountered by learners during training, it was important to determine the types of errors made by the groups. Table 9 shows the number of tone-only errors expressed as a proportion of the total errors within each training session and the generalisation test.

Table 9. *Proportion of tone-only errors made (out of total errors made) by all learner groups in sessions 1 to 5 and the generalisation test*

Group	Tone-only errors (%)					
	Session 1	Session 2	Session 3	Session 4	Session 5	Generalisation
English	40	60	66	73	79	82
Mandarin	58	87	84	84	87	87
Vietnamese	62	86	86	90	93	92

In session 1, a large proportion of errors made by all groups were segment-only or segment and tone errors. By session 5, all groups had managed to learn most of the segmental contrasts and the majority of errors made were tone errors. Despite the drop in performance in the generalisation test (see Figure 6), the proportion of tone-only errors to total errors in this task was almost identical to that of the session 5 training task. Like the tonal groups, the AusE group were making mostly tone-only errors by the end of training. However, as observed earlier, they still made more errors than the tonal groups, which suggests they have not been able to show the same level of sensitivity to tone differences as the Mandarin and Vietnamese tonal groups have.

3.2.2 Confusion matrices

Confusion matrices were also created for tone training and generalisation. Unlike the tone identification task, learners in the training and generalisation tasks can misidentify a target checked tone as an unchecked tone and vice-versa, since learners were provided with all 16 response options during the test. Confusions between checked and unchecked tones occurred less frequently than confusions between checked tones and confusions between unchecked tones. For the sake of brevity, only confusions from training sessions 1 and 5 will be discussed.

Table 10. *Tone training session 1 confusion matrices*

Group	Stimulus tone	Selected tone label					
		T1 (33)	T2 (11)	T3 (41)	T4 (51)	T5 (55)	T6 (41)
AusE	T1 (33)	**24**	14	25	22	7	7
	T2 (11)	10	**43**	22	16	4	5
	T3 (41)	19	21	**30**	21	3	5
	T4 (51)	23	13	25	**26**	8	6
	T5 (55)	10	2	8	6	**53**	22
	T6 (41)	9	8	12	11	15	**45**
Mandarin	T1 (33)	**52**	12	11	17	3	4
	T2 (11)	17	**44**	14	20	2	3
	T3 (41)	10	19	**38**	28	2	3
	T4 (51)	15	16	30	**35**	2	3
	T5 (55)	0	3	8	5	**73**	11
	T6 (41)	3	1	5	5	28	**58**
Vietnamese	T1 (33)	**31**	11	19	31	4	4
	T2 (11)	17	**40**	20	19	2	2
	T3 (41)	21	23	**23**	27	2	5
	T4 (51)	28	12	19	**35**	2	4
	T5 (55)	7	2	10	4	**56**	20
	T6 (41)	4	4	11	4	20	**57**

In session 1 of the tone training task, accuracy in identifying the target word's tone was at or below chance for most tones. For all groups, identification accuracy was highest for the checked tones. The AusE and Mandarin groups showed higher identification accuracy for T5 over T6, while accuracy for both tones was similar for the Vietnamese group. All groups mostly confused the checked target with the other checked tone. Likewise, when the target tone was unchecked, the learners' confusions comprised mostly of the other unchecked tones. The confusions for some unchecked stimulus tones were spread across all unchecked tones, given the low identification scores at session 1. Some confusion patterns observed in the tone identification task were also reflected in the tone training task. For instance, the AusE group were more likely to confuse the mid level T1 as a falling tone than a low level tone. Similarly, the falling T4 was confused with the other falling tone and the mid level T1, but less often as a low level tone. T2 was often confused as T3, which has a lower starting pitch height and smaller pitch excursion than the other falling tone, T4. The Mandarin group's most prominent confusion patterns were for the falling T3 and T4, both of which were often confused for one another. Though their pretest tone identification performance was high for both tones, it seems that the group struggled to differentiate the two in tone word learning. This suggests that the Mandarin native falling category may be interfering with the group's ability to distinguish two nonnative falling tones, even if one of the nonnative tones is similar to the native tone. The Vietnamese group had difficulties correctly identifying T3, identifying it more often as T4. However, unlike the Mandarin group, they confused T4 as T1 and vice-versa, indicating that they struggled to attune to the differences between the falling tone (51) and the mid level tone (33).

Table 11. *Tone training session 5 confusion matrices*

Group	Stimulus tone	Selected tone label					
		T1 (33)	T2 (11)	T3 (41)	T4 (51)	T5 (55)	T6 (41)
AusE	T1 (33)	**48**	5	21	25	0	1
	T2 (11)	5	**63**	23	7	1	1
	T3 (41)	17	12	**45**	24	1	1
	T4 (51)	27	7	18	**45**	0	3
	T5 (55)	2	1	1	0	**83**	14
	T6 (41)	1	2	1	1	14	**82**
Mandarin	T1 (33)	**77**	12	4	6	1	1
	T2 (11)	8	**74**	12	5	0	0
	T3 (41)	3	9	**60**	27	0	0
	T4 (51)	4	3	26	**65**	1	1
	T5 (55)	0	0	2	0	**90**	8
	T6 (41)	0	0	1	0	7	**91**
Vietnamese	T1 (33)	**65**	4	15	16	0	0
	T2 (11)	10	**62**	20	9	0	0
	T3 (41)	15	19	**44**	23	0	0
	T4 (51)	16	4	18	**61**	0	0
	T5 (55)	0	0	0	0	**86**	14
	T6 (41)	0	0	0	0	5	**95**

By the end of tone training, all learner groups improved in their tone identification accuracy, particularly for the checked tones. Confusions between checked and unchecked tones reduced substantially; the proportion of confusions between the two categories of tones was no greater than 3% across groups. Some general patterns can be gleaned from the confusion matrix data. Level tones were identified with higher accuracy than the falling tones—the AusE group showed higher identification accuracy for T2, the Vietnamese group for T1, and the Mandarin group for both. Across all groups, T3 was the least accurately identified tone. The AusE and

Vietnamese groups showed very similar confusion patterns to each other, some of which remain from session 1. Both groups tended to confuse T1 and T4 with the other unchecked tones but not T2 (however, the AusE group did confuse T4 with T1 most often), and T2 was often confused as T3. Furthermore, the confusions for T3 in both groups were spread across tones, with a slight bias towards misidentifying the stimulus as the other falling tone, T4. Though their accuracy was higher than the other groups, the Mandarin group showed a more obvious bias towards misidentifying T3 as T4 and also predominantly misidentified T4 as T3. It was clearer in session 5 that the Mandarin group was confusing the two falling tones, perhaps as a result of only having one falling tone category in their native language, yet their scores for these tones remained higher than the other groups. As posited earlier, the difficulty in identifying T3 might be due to the tone's similarity to T4, with its major differences including a lower starting pitch and a smaller pitch excursion. This may make T3 not only confusable as T4, but also confusable as a level tone, hence the spread in confusions for the AusE and Vietnamese groups. The Mandarin group appears to be more sensitive to differences between a level and contour tone, with their main weakness being the differentiation of two falling tones.

Table 12. *Generalisation test confusion matrices*

Group	Stimulus tone	Selected tone label					
		T1 (33)	T2 (11)	T3 (41)	T4 (51)	T5 (55)	T6 (41)
AusE	T1 (33)	**34**	15	24	25	1	1
	T2 (11)	8	**41**	39	11	0	2
	T3 (41)	17	22	**38**	20	2	1
	T4 (51)	22	13	28	**36**	0	1
	T5 (55)	2	2	1	0	**78**	17
	T6 (41)	3	0	0	2	20	**74**
Mandarin	T1 (33)	**75**	14	4	6	0	0
	T2 (11)	4	**45**	34	16	0	1
	T3 (41)	3	15	**54**	26	1	1
	T4 (51)	5	7	29	**58**	1	0
	T5 (55)	0	1	1	0	**91**	7
	T6 (41)	0	1	1	0	8	**90**
Vietnamese	T1 (33)	**63**	6	13	18	0	1
	T2 (11)	11	**45**	32	12	0	1
	T3 (41)	11	22	**47**	19	0	0
	T4 (51)	20	7	20	**52**	0	1
	T5 (55)	0	0	0	0	**88**	11
	T6 (41)	0	0	0	0	10	**89**

Even when listening to productions from a novel talker, confusion patterns remained similar to those in session 5. The checked tones were still identified with high accuracy by all groups, and the Mandarin and Vietnamese groups identified T1 with relatively high accuracy. All groups showed difficulty in identifying T2 accurately, often confusing it as T3. Confusions for T3 were spread across the unchecked tones, though the Mandarin group mostly confused T3 as T4 (and mostly confused T4 as T3). Like in session 5, the AusE and Vietnamese groups confused T1 and T4 with the other unchecked tones but were better at differentiating both target tones from

T2. In sum, all groups may have been perceiving the low level T2 as a falling tone in the generalisation test. The Vietnamese group appears to have performed more similarly to the AusE group in tone word learning (though the Mandarin and Vietnamese groups bear some similarities in confusion patterns in the tone identification task), as the AusE and Vietnamese groups showed a similar misidentification spread for the mid level T1 and falling T4. Conversely, the Mandarin group's confusion patterns were narrowed down to difficulties with distinguishing between the falling tones T3 and T4.

Chapter 4: Discussion

The present study investigated nonnative tone learning in tonal and nontonal language groups. Nontonal English, tonal Mandarin, and tonal Vietnamese listeners completed a five-session Hakka tone training program consisting of tone identification, tone word learning, and generalisation tasks. Participants were invited to return three weeks after the cessation of training to complete a tone identification retention test, and the majority did so. The study addressed three research questions. The first concerned whether native tone language experience facilitated nonnative tone learning. The second sought to determine whether having a more complex (denser) native tone system led to better tone learning outcomes. The third and final question was concerned with whether native tonal and intonational categories can be used to predict the successful learning of nonnative tones. It was hypothesised that both tonal groups would outperform the nontonal English group before and after tone training. Regarding the second question, if a more complex native tone system provides additional tone learning benefits, the Vietnamese group would outperform the Mandarin group. Alternatively, successful tone learning can instead be predicted by observing the similarities between the nonnative tones and the learner's native tonal or intonational categories. These predictions, in line with the principles of the Perceptual Assimilation Model, can provide insight on which tones and tone contrasts the learner groups may find easier or more difficult to distinguish. The results lend support to the tone experience and native categories hypotheses, but tonal complexity did not have a cumulative effect on nonnative tone learning.

4.1. The effect of language background on tone learning performance

All learner groups, regardless of prior tone language experience, showed significant improvement in tone identification and tone word learning following five sessions of tone training. Not only were the listeners able to differentiate nonlexical tone contrasts after the training program, they also learned to associate pictures with sounds and to differentiate these novel picture-sound pairings using variations in pitch. Three weeks after the cessation of training, the Mandarin and Vietnamese groups showed no decline in tone identification performance, indicating that the gains of tone training were maintained. The findings are consistent with evidence that naïve adult listeners from tonal and nontonal language backgrounds are able to learn nonnative tones via training (Cooper & Wang, 2012; Francis et al., 2008; X. Wang, 2013; Wayland & Li, 2008; Wong & Perrachione, 2007). Further, by the end of training, all learner groups predominantly made errors during the sound-to-image testing phase by selecting the incorrect tone. That is, the participants were able to learn segmental differences in the training stimuli, and results (at least in session 5 and the generalisation test) were not impacted severely by the misidentification of segments. The learners showed improvements in all sessions except between sessions 4 and 5, which may suggest that by the fifth session the groups had started to asymptote. However, we may have observed a different pattern if the learners were separated into good and poor learner subgroups on the basis of individual differences in pretraining tone identification abilities. In one study, poor learners showed no significant improvement between sessions 4 to 5 and sessions 5 to 6, whereas the good learners continued to show improvement between these sessions (Chandrasekaran et al., 2010). Despite the improvement in performance after training, none of the groups in the current study reached a level of attainment comparable to results in other studies (Antoniou & Wong, 2016; Wong & Perrachione, 2007), perhaps due to the learners only being separated by native language and not learner aptitude. Additionally, the Hakka tone contrasts used here may be more difficult to discern than the Mandarin contrasts used in these studies. In future, it would be worth investigating differences in tone learning performance across language groups and across learners of low and high aptitudes with an expanded set of tone languages.

Both Mandarin and Vietnamese groups outperformed the AusE group in the tone identification and generalisation results, but only the Mandarin group outperformed the AusE group in tone word learning. It would appear that the tonal groups benefitted from an initial boost in performance in the tone identification task, and the Mandarin group in the tone word learning task, leading to higher identification accuracy in the final session of training. Note that the rate of improvement remained constant between all groups, as there was no interaction effect between language and test/session in the tasks. Despite the Vietnamese group showing no significant differences from the AusE group, results showed that tonal language speakers were also better than nontonal language speakers at perceiving nonnative speech produced by a new talker (generalisation test). Native tone language listeners appear to possess an advantage over nontonal language listeners in nonnative tone identification. Similarly, the Mandarin listeners were facilitated by their native language when learning nonnative tone words, yet the Vietnamese listeners did not differ from the other learner groups in this task. These observations can be explained as follows. Tone identification is a lower level task because the nonce stimuli do not carry semantic information. This explanation is consistent with the reports in some studies that musical experience was found to be a better predictor of successful tone identification than tone experience, as musicians rely more on lower-level information when processing musical pitch (Cooper & Wang, 2012; Zhao & Kuhl, 2015). Conversely, tone word learning, which involves the use of lexical tone to identify the meanings of words, requires processing at a higher level and this will benefit from long-term tone language experience. The findings in the current study contrast with prior evidence showing that prior tone language experience does not contribute to higher tone identification and tone word learning performance (Francis et al., 2008; X. Wang, 2013).

Though musical experience did not have any effect in most of the tasks in the present study, it did affect tone identification performance between the two tonal learner groups. A number of participants in the current study, predominantly listeners in the Mandarin group, had several years of musical experience. To test whether musical experience showed a significant effect on tone learning, scores for each task were submitted to ANCOVAs, with years of musical experience entered as a covariate. Like tone experience, musical experience also grants listeners domain-general abilities to utilise tonal information at lower levels of processing, but having experience in both linguistic and musical pitch does not provide an additive advantage for nonnative tone learning (Cooper & Wang, 2012, 2013). Interestingly, musical experience did not have an effect on tone identification and tone word learning performance, except for the comparison between the Mandarin and Vietnamese groups at post-test and retention test. The Mandarin group significantly outperformed the Vietnamese group, although this advantage was not maintained when years of musical experience was included as a covariate. This suggests that musical experience may have had some effect on tone identification performance in this group of native tone language listeners. However, musical experience was not the focus of this study (i.e., we did not compare Mandarin musicians vs. nonmusicians), and further research is required to understand how musical experience interacts with tone language experience when learning nonnative tones.

For the most part, the results of this study have supported the hypothesis that native tone language experience provides an advantage for nonnative tone perception. The Mandarin listeners were the only group to show this advantage for all tasks, while the Vietnamese group showed no tone experience benefits for tone word learning. It had been expected that the

Vietnamese group would rely on higher-level phonemic information in order to learn the nonnative tone words, yet they did not differ significantly from the AusE or Mandarin groups. Further, the Mandarin listeners outperformed the Vietnamese listeners in the tone identification task, contradicting prior evidence of tonal complexity facilitating nonnative tone perception. The role of tonal complexity on tone learning is discussed in detail in the following section.

4.2. Tonal complexity and nonnative tone learning

Tonal complexity has been defined as the number of tones within a tone system. Previous studies have shown that tonal complexity facilitates nonnative tone perception such that tonal listeners with a denser native tone inventory are more sensitive to pitch variations than listeners with a sparser tone inventory (Tong & Tang, 2016; Zheng et al., 2012). Other findings have shown the opposite effect, whereby having a more complex tone system causes native categories to act as perceptual magnets, thus interfering with discrimination of nonnative tones with similar properties (Chiao et al., 2011). Granted, there are relatively few tone training studies that compare the performance of different tone language groups, and so the effect of tonal complexity on tone learning is inconsistent. Here, it was hypothesised that if tonal complexity predicts tone learning success, the Vietnamese group would outperform the Mandarin group, as Vietnamese has five tones and two checked tones and Mandarin only has four tones. The results force us to reject this hypothesis; we observed that not only did the Mandarin group outperform the AusE group in tone training, they also outperformed the Vietnamese group in the tone identification task. Therefore, the current study provided strong evidence that tonal complexity does not facilitate nonnative tone learning. If tonal complexity does not contribute to successful tone learning, then there must be another approach to making predictions on learning performance for listeners in different tone language groups. Alternatively, the current definition of tonal complexity is insufficient and may require revision.

Earlier in Section 1.1.1, we raised the problems associated with definitions of tonal complexity that only consider the density of a tone system, because this ignores other relevant features of tone. It is possible that past work that has found support for the tone density explanation has obscured the underlying mechanisms responsible for between-group differences and misattributed the reason for the learning patterns observed to tonal complexity. The tone language groups tested in the vast majority of tone perception studies (including the current study) generally belong to what Maddieson et al. (2011) classify as complex tone languages: languages with a tone system containing at least four tones. Note, for example, that the tonal learner groups in past studies are native speakers of a language with a complex tone system, and that in both cases, the successful learners were native speakers of Cantonese (Tong & Tang, 2016; Zheng et al., 2012). Learning advantages may be observed for listeners with a denser tone system if performance was compared between a complex tone language group and moderately complex (with three tones) or simple (with two tones) tone language groups. Outside of the tone perception literature, other definitions of tonal complexity have focused instead on individual tones rather than the entire tone inventory. For instance, Zhang's (2004) definition takes into account the following features of a tone: the number of pitch targets, the excursion of the contour between pitch targets, and the contour's direction. Perhaps the tonal complexity of a tone language is better determined through examining the complexity of tones within the system, and testing this definition could provide insight on ways to compare performance between tonal learner groups.

In sum, more research is needed on the various methods of determining the complexity of a tone or tone system before we can establish a more comprehensive definition of tonal complexity. An updated definition may provide more reliable predictions for how certain tone language groups may perceive a nonnative tone, or future results may show that no such predictions can be made using tonal complexity alone. In this case, we could instead make predictions by observing the similarity between nonnative tones and the listener's native tonal or intonational categories, as we have done for the present study. This approach reflects prior studies which have tested the predictions of PAM on tone perception (Hallé et al., 2004; Reid et al., 2015; So & Best, 2010, 2014). In the next section, we discuss how the Mandarin group's superior performance over the Vietnamese group in tone identification might be attributed to Hakka tones mapping more closely onto Mandarin rather than Vietnamese categories.

4.3. Similarities between nonnative tones and native tonal/intonational categories

Past studies have shown that native tonal or intonational categories can affect the perception of nonnative tones. Nonnative tones may be mapped onto existing categories, which can facilitate learning if no other nonnative tone occupies this space; but when multiple tones are heard as the same native category, patterns of assimilation become more complicated. Speech perception models such as PAM (Best, 1995; So & Best, 2010) have been adapted to examine the similarities between nonnative tones and the L1 phonological space in order to predict patterns of discrimination. Cue weighting also affects a listener's sensitivity to pitch dimensions. For instance, a nontonal language speaker may place more weight on perceiving pitch height, while a tonal language speaker would place more weight on pitch direction (J. T. Gandour & Harshman, 1978; Guion & Pederson, 2007; Xu et al., 2006), though some speakers of tonal languages such as Cantonese give more weight to both dimensions than other language speakers (Francis et al., 2008). There are also tones and tone pairs that are universally easier to perceive, such as rising tones, which are more perceptually salient than falling tones (Krishnan et al., 2010a, 2004); and the DynamicDynamic tone pair, which comprises of rising and falling tones (Burnham, Kasisopa, et al., 2015). However, these tones and tone pairs are not present in the Hakka tone stimuli.

To further examine the relation between nonnative tones and the learners' native languages, we made predictions on the language groups' tone identification patterns based on PAM principles, as well as the learner's weightings towards perceptual cues in pitch. All groups were predicted to find the T1–T2 (33 vs. 11), T2–T4 (11 vs. 51), and T5–T6 (55 vs. 41) contrasts easy to distinguish, while the T3–T4 (41 vs. 51) contrast was predicted to be the most difficult to distinguish due to the tones' phonetic similarity. The Mandarin learners were predicted to identify all Hakka tones with high accuracy due to how well the tones map onto native tone categories, but would struggle with the T3–T4 contrast since they only have one native falling tone category (51). Since Hakka T4 is most similar to Mandarin T4, we expected the Mandarin listeners to show Single Category or Category Goodness assimilation patterns by misidentifying T3 as T4. While Hakka tones do not map as consistently onto the Vietnamese tone system, the Vietnamese learners were expected to identify T1, T2, and T5 with relatively high accuracy. They were also expected to show poorer accuracy for T3 and T4 since neither tone maps perfectly onto the Vietnamese falling tone (21). AusE listeners, who give more weight to the pitch height dimension when perceiving tone, were also predicted to show higher identification accuracy for the level tones (T1, T2, and T5). Differences between falling intonation contours exist in English for *Statement* and *Exclamation* intonation patterns, where the *Exclamation* contour begins at a higher pitch level than *Statement*, much like the falling tones in Hakka. But it was not clear if the AusE listeners in this study would have been sufficiently sensitive to this slight variation in pitch in the Hakka tones, so we predicted that they would also struggle to distinguish the T3–T4 contrast, assimilating both nonnative tones into one of the native falling categories. We created confusion matrices to observe the learners' tone identification patterns for each tone across the individual tasks, and also for when the tones were paired with a particular visual distractor tone. Some of these predictions align with the tone identification and tone word learning results, while observations were also made which can explain the effect of native language categories on nonnative tone learning.

In the tone identification pretest, all learner groups were already choosing the intended category with over 50% accuracy, with the Mandarin group showing high identification accuracy for all tones ($\geq$ 80%). These scores improved following the five-session training program, and the tonal groups' performance did not decline three weeks following training. Several observations can be made from the pretest, post-test, and retention confusions. As predicted, all groups across all tests identified the level tones (T1, T2, T5) more consistently than the falling tones (T3, T4, T6). Of the unchecked level tones, the tonal group identified T1 most accurately, whereas the AusE group identified T2 with the highest accuracy. From looking at the full confusion matrix alone, it appears that these Hakka tones may have been mapped onto native categories, resulting in higher tone identification accuracy. Hakka T1 (33) could have mapped onto Mandarin T3 (214), Vietnamese A1 (44), and the AusE *Flat-pitch* intonation contour; Hakka T2 (11) to Mandarin T3 once again, Vietnamese B2 (212) or C1 (214), and the general low level intonation contour in AusE; and Hakka T5 (55) to Mandarin T1 (55), Vietnamese A1 (44), and AusE's high level intonation contours. Possible mappings of the Hakka falling tones to native categories were outlined in the previous paragraph. As explained below, observing the tone identification performance for tone contrasts reveals a more complicated relationship between nonnative tones and native tone and intonation categories.

Cue weighting also plays a role in nonnative tone perception. AusE listeners place greater weighting on pitch height when perceiving tone, whereas the Mandarin group attends more to pitch direction as their native language contains rising, dipping and falling tones, the Hakka tone system consists of only falling tones. Southern Vietnamese speakers rely less on voice quality compared to Northern Vietnamese speakers and attend to pitch cues instead, though the weighting towards either pitch height or contour is unspecified (Brunelle, 2009). Considering the slight differences in starting pitch height between T3 and T4, it would make sense for the groups to identify T1 and T2 with greater accuracy. Out of the two falling tones, T3 was identified less consistently than T4 by all learner groups, but by post-test and retention test, only the AusE group identified both falling tones below 70% accuracy. Although the Mandarin listeners performed worse on identifying the falling tones, they still showed higher identification accuracy than the other groups for all tones. We had predicted that T4, the tone with a higher pitch height at onset and a larger pitch excursion, would be more perceptually salient than T3. Like the mid level T1, we posited that T3 would be more confusable as T4 since T3 fell within the two pitch height extremes (i.e., low and high). However, upon looking at the identification accuracy for tones in certain tone pair contexts, we found that T3 was identified with higher accuracy than T4 in the T3–T4 contrast, suggesting T4 was actually more confusable as T3. In fact, the lower identification score for T3 was due to the spread of misidentifications across all the other tones. The tone contrasts in which T3 showed the lowest identification score varied by language group. These tone contrast identification patterns are discussed below.

As predicted, all learner groups consistently distinguished the T1–T2 and T2–T4 contrasts, but identification accuracy for T1–T2 was more noticeable at post-test for the AusE group. This is possibly due to the tones in these contrasts starting at different pitch heights (mid vs. low, and low vs. mid-high), thus making differences between the tones more noticeable earlier on. In past studies, English listeners have been shown to be more sensitive to earlier pitch differences, contrasting Mandarin listeners' higher sensitivity to later pitch differences (Kaan et al., 2007). Since T1–T2 was identified with high accuracy by the Mandarin learners, this meant that both

T1 and T2 could not have been heard as the same native category, Mandarin T3 (214). Perhaps Hakka T1 was mapped onto the neutral Mandarin tone, instead. As for the AusE group's later improvement in distinguishing the T1–T2 contrast, this result parallels past findings showing that English listeners can initially distinguish between low and high level tones only, but can learn to distinguish a third level tone after receiving tone training (Francis et al., 2008). Compared to the low level T2, the AusE listeners identified the mid level T1 with lower accuracy and also showed less improvement in identifying the tone at post-test, which indicates that the group may still be learning to attend to fine-grained pitch differences. Sensitivity to early pitch differences may also explain the lower identification accuracy for T6 compared to T5 in the T5–T6 contrast, as the pitch height of both tones begin at a similar level. This pattern occurred for both the AusE and Mandarin groups, while identification scores for T5 and T6 were similar for the Vietnamese group. Nevertheless, accuracy for the checked tones was higher overall since listeners only had to differentiate between two visual category response options whenever a checked tone was presented as the stimulus.

It is interesting to note that all groups distinguished T2 from T1 and T4 with high accuracy, but T2 was not as easily distinguishable from T3, especially for the Vietnamese group. For the AusE group, whose average identification scores were lower than the other groups', T2–T3 was one of their more consistently identified contrasts. T2 is located on the lowest pitch level and is furthest away from the other tones at tone onset, which may make it more distinguishable from the level T1 and the high falling T4. However, it appears that T3 sounds similar to T2 for the learners, leading to a decrease in identification accuracy. T2 (11) and T3 (41) may not have mapped onto Vietnamese A2 (21), B2 (212) or C1 (214), hence the poorer identification accuracy for this group. For the Mandarin group, T2 and T3 were identified with similar accuracy (which was still higher than the other learners' scores), suggesting that they are both equally difficult to distinguish. In the AusE and Vietnamese groups, T3 was clearly more confusable as T2, which implies that T3 was being heard as a low level tone. Both groups found additional difficulties in perceiving tone contrasts involving T1, T3 and T4; conversely, the Mandarin group excelled in distinguishing the T1–T3 and T1–T4 contrasts. The AusE and Vietnamese groups may have heard T3 and T4 as mid level tones, as the both groups showed poor identification accuracy for the falling tones when the visual distractor tone was T1. This confusion was one-sided for the Vietnamese group, as identification accuracy for T1 remained consistent when T3 and T4 were the distractors, but the AusE learners did show a slight decrease in accuracy for T1 in these contexts. Examining identification patterns in tone contrasts has revealed assimilation patterns that differ from those predicted in the hypothesis. For the Vietnamese and AusE groups, Hakka T3 may have been mapped to Vietnamese B2 (212) or C1 (214), and the low level AusE intonation contour. T3 and T4 may have been mapped to Vietnamese A1 (44) and the *Flat-pitch* AusE intonation contour. The AusE listeners may have also mapped T1 as a *Statement* or *Exclamation* intonation contour, so it is possible that the AusE group showed Single Category assimilation for the T1–T3 and T1–T4 contrasts. The Vietnamese group, on the other hand, may have shown Category Goodness assimilation for these contrasts, with T1 being the ideal nonnative exemplar of the native category.

With this in mind, we can then analyse one of the more complicated identification patterns. As mentioned earlier, all learner groups showed varying levels of difficulty in distinguishing between the tones in the T3–T4 contrast. This finding aligned with our predictions, yet our assumption that T3 would be more confusable than T4 was incorrect: identification accuracy for both tones were similar, but the AusE group showed higher identification accuracy for T3 than T4. It is possible that the AusE learners showed higher identification accuracy for T3 because it is a more acceptable exemplar of the native falling intonation category in English (*Statement*), while T4 is the poorer exemplar. For the AusE and Vietnamese groups, identification accuracy for T3 was still higher in the T3–T4 contrast than in other T3 tone contrasts. This indicates that both groups may have perceived T3 as a low or mid level tone, leading to lower identification scores when a level tone was a visual distractor and slightly higher scores when T4 was the visual distractor. The Mandarin group did not show this pattern, though T3–T4 was the contrast that they had the most difficulty distinguishing. Across the pretest, post-test and retention test, their identification accuracy for T3 and T4 decreased to the point where the Vietnamese group began to differentiate T3 and T4 with higher consistency. Initially, the Mandarin listeners appeared to have been able to tease apart the fine-grained pitch variations between T3 and T4, but prolonged exposure to the tones may have led them to assimilate the tones to a single category, thus decreasing the identification accuracy for both tones. In sum, there were language-general, as well as language-specific, patterns of tone identification: among all learners, level tones were easier to identify than falling tones, and T1–T2 and T2–T4 were the most easily distinguishable contrasts. While all groups showed some difficulty in differentiating certain tone contrasts, the AusE group had the most trouble with contrasts containing T1 and the falling tones, the Vietnamese group struggled to distinguish T2–T3 since the tones did not map closely onto native categories, and the Mandarin group's identification accuracy for T3–T4 continued to decline from pretest through to retention test. Surprisingly, the Vietnamese group's identification patterns matched more closely those of the AusE group's, but also showed similarities to the Mandarin group for certain identification patterns, such as their high accuracy for T1.

Tone identification patterns between the groups were similar in the tone training task (tone word identification), even though there were 16 response options in the training task as opposed to two for tone identification. All groups performed at or below chance in the first session, and the errors made by the learners were predominantly segmental-only or both tonal and segmental. By the final session, all learners improved in their tone identifications and were able to learn the segments of the tone words. Despite this, tone confusion patterns were similar between sessions 1 and 5. The checked tones were identified with the highest accuracy, followed by level tones. T3 was still the least accurately identified tone, with confusions spread across the other unchecked tones for the AusE and Vietnamese groups. As seen in their tone confusions, Mandarin listeners often misidentified T3 as T4 and T4 as T3, also showing a Single Category or Category Goodness assimilation pattern in the training task. The AusE and Vietnamese groups once again showed similar confusion patterns to each other: they confused T1 as the falling tones and the falling tones as T1, but could distinguish T1 and T2 well. Both groups also confused T2 as T3, but not as T4, which may indicate that the lower starting pitch height in T3 causes the listeners to confuse the low level T2 as T3. Overall, the confusion patterns in the tone word learning task reflected those observed in the tone identification task.

Most of the identification patterns remained in the generalisation test, but the decline in performance also led to less consistent confusion patterns. One pattern which contradicts the findings in the tone identification task is the decrease in T2 accuracy in the generalisation test. T2 was still confused as T3, but this was observed in all learner groups' responses instead of just the AusE and Vietnamese groups. However, this confusion pattern appears to have been caused by the pitch values that were superimposed onto the talker's tone stimuli (see Table 1). The T2 tone values for speaker 3, who produced the generalisation stimuli, are higher at pitch onset and offset than the other speakers' tone values. Also, for a low level tone, the pitch excursion from onset to offset is larger, which might have made the listeners misidentify the target tone more often as T3. Apart from these findings, the generalisation results did not differ much from tone word learning.

4.4. Limitations and future directions

The current study tested native speakers of Australian English, Mandarin and Vietnamese. Although we were able to recruit AusE speakers with no tone language experience (note that one participant spoke Serbian at home and another was studying Japanese, both of which are pitch-accented languages) and primarily Southern Vietnamese speakers, we encountered an issue when recruiting Mandarin participants. Firstly, none of the Mandarin participants reported prior experience with Hakka Chinese, but exposure to the dialect is difficult to determine, especially for participants who have lived extended periods of time in areas where Hakka is commonly spoken, such as in the southern provinces of China, Hong Kong, Taiwan, and in Chinese-speaking communities in Southeast Asia. Secondly, around half of the Mandarin participants reported proficiency in another non-Hakka Chinese dialect, though almost all reported dominance in Mandarin. This can be problematic, as knowledge of another tone system may provide additional benefits for tone language speakers and may affect how nonnative tones map onto established tonal categories.

Improvements could be made for certain aspects of the study's design. Firstly, responses for the tone identification test trials were limited to three seconds, meaning the trial was considered a missed trial if listeners did not respond in time. We observed a few missed trials in the output for a number of participants, particularly the first few trials of the test phase at pretest. One participant had to be excluded from the study due to an excessive number of missed trials. Past studies have also set a time limit for test trials (So & Best, 2010), but the interstimulus interval has sometimes been extended to as long as 10 seconds (Cooper & Wang, 2012). Tasks in other studies, though not necessarily tone identification, have been self-paced, thus ensuring a response is made for every trial (Francis et al., 2008; Wu et al., 2014). Another issue to consider is the training protocol: this study consisted of five training sessions, but additional sessions may have shown further improvement for the learners. Though there was no significant difference in performance between sessions 4 and 5, this does not preclude the possibility that more progress could be have been made in subsequent sessions. Wong and Perrachione (2007) tested participants until they met a certain training criterion—over 95% for two consecutive days of training or less than 5% improvement in accuracy for four consecutive days—and found that more successful and less successful learners reached asymptotic performance in a similar number of sessions (7.22 vs. 9.38). Perhaps a seven- or eight-day training program may be sufficient for observing such learning performance.

This study has implications on future research involving nonnative tone perception and tonal complexity. Future studies may continue to adapt speech perception models such as PAM to tone training research by incorporating other experimental tasks such as AXB discrimination and categorisation. Ideally, an entirely new framework should be established that is tailored to account for the specific effects that have been observed for tone perception (and that differ from segmental perception). One step towards this direction is to evaluate the ways in which listeners from different language and musical backgrounds perceive qualities of tone. These include, but are not limited to, universal aspects of tone perception such as the saliency of rising pitch contours and the discriminability of tone pairs (Burnham, Kasisopa, et al., 2015), and the weighting of perceptual pitch cues such as height, direction, and voice quality (Brunelle, 2009; J. T. Gandour & Harshman, 1978; Guion & Pederson, 2007). Another concept that requires more discussion is tonal complexity, which we argue should be conceptualised in a more comprehensive way than currently used definitions. Tonal complexity can still comprise of elements such as the density of a tone system (Tong & Tang, 2016; Zheng et al., 2012), but should also account for the complexity of individual tones within the tone inventory (Zhang, 2004). An increased understanding on how tones interact within the system and their relation to nonnative tones would allow for a more holistic approach to tone perception research.

Chapter 5: Conclusion

This thesis investigated the effect of native tone language experience on nonnative tone learning, and whether tonal complexity, or the density of a tone system, grants additional learning benefits. We trained listeners from different language backgrounds—nontonal Australian English, tonal Mandarin, and tonal Southern Vietnamese—to learn Hakka Chinese tones. Based on prior observations in the literature, we predicted that the tonal language listeners would outperform the AusE listeners. Under the tone complexity hypothesis, speakers of Vietnamese, a language with five tones and two checked tones, would outperform speakers of Mandarin, a language with four tones. An alternative approach to predicting nonnative learning success is to consider the interactions between nonnative tones and native tonal and intonational categories, similar to studies which have tested the predictions of the Perceptual Assimilation Model.

Over five training sessions, listeners completed tone identification and tone word learning tasks. All learners exhibited improvement in tone processing, as measured by accuracy in the tone identification and word learning tasks, after completing the fifth and final training session. The Mandarin and Vietnamese groups who returned to complete the tone identification retention test showed no decline in performance three weeks following the cessation of training. Findings showed a significant effect of tone experience on tone identification and tone word learning, with Mandarin listeners outperforming the AusE listeners in all tasks. Vietnamese listeners also outperformed AusE listeners in tone identification and generalisation but did not differ from the other two groups in tone word learning. The Mandarin group outperformed the Vietnamese group in tone identification, indicating that tonal complexity (density) does not predict successful tone learning. Tone identification patterns revealed that all learners identified level tones more consistently than falling tones, and that the T1–T2 and T2–T4 contrasts were among the easiest to distinguish. Some confusion patterns were specific to learner groups, though Vietnamese learners showed similar identification patterns to AusE learners, as well as Mandarin learners. AusE listeners, who showed more sensitivity to pitch height, struggled to distinguish mid level T1 from the falling T3 and T4. Although the Mandarin listeners identified all tones with high accuracy, their accuracy was lowest for the T3–T4 contrast as a result of both falling tones interacting with the single falling tone of Mandarin. Vietnamese listeners had difficulties distinguishing the T2–T3 contrast, perhaps due to the nonnative tones mapping poorly onto their native tone categories.

The findings have implications for theoretical models of nonnative tone perception, and may lay the groundwork for the development of a framework that will specifically account for the perception of lexical tone rather than segments. Such a model would incorporate concepts such as cue weighting and universal aspects of tone perception, as well as tonal complexity, which, as it is currently defined, accounts for only one aspect of a tone system: density. We propose that analysing the complexity of individual tones within the tone system is just as important. Re-evaluating the definition of tonal complexity is a future direction that could lead to more informed predictions on nonnative tone learning for speakers of tonal and nontonal languages.

References

Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, *21*(4), 319–325. doi:10.1177/002383097802100406

Alexander, J., & Wang, Y. (2016). Cross-language lexical-tone identification. *Proceedings of the 5th International Symposium on Tonal Aspects of Languages (TAL 2016)*, 28–32. doi:10.21437/TAL.2016-6

Antoniou, M., & Chin, J. L. L. (2018). What can lexical tone training studies in adults tell us about tone processing in children? *Frontiers in Psychology*, *9*(1), 1–12. doi:10.3389/fpsyg.2018.00001

Antoniou, M., & Wong, P. C. M. (2015). Poor phonetic perceivers are affected by cognitive load when resolving talker variability. *The Journal of the Acoustical Society of America*, *138*(2), 571–574. doi:10.1121/1.4923362

Antoniou, M., & Wong, P. C. M. (2016). Varying irrelevant phonetic features hinders learning of the feature being trained. *The Journal of the Acoustical Society of America*, *139*(1), 271–278. doi:10.1121/1.4939736

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, *125*(2), 238–251. doi:10.1093/brain/awf028

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

Beckman, M. E. (2012). *Stress and non-stress accent*. Dordrecht: Walter de Gruyter.

Beckman, M. E., & Hirschberg, J. (1994). *The ToBI annotation conventions.* Online MS. Available at /http://www.ling.ohiostate.edu/~tobi/ame_tobi/annotation_conventions.htmlS; Last accessed November 1, 2018.

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance, 32*(1), 97–103. doi:10.1037/0096-1523.32.1.97

Best, C. T. (forthcoming). The diversity of tone languages and tonality in non-tone languages: How well does perceptual research reflect this diversity? *Frontiers in Psychology*.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Retrieved from http://www.haskins.yale.edu/Reprints/HL0996.pdf

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345–360. doi:10.1037/0096-1523.14.3.345

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementaries. In Murray J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). Amsterdam: John Benjamins.

Best, C. T., Tyler, M. D., Goldstein, L. M., & Nam, H. (2016). Articulating what infants attune to in native speech. *Ecological Psychology*, *28*(4), 216–261. doi:10.1080/10407413.2016.1230372

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PloS One*, *8*(4), e60676. doi:10.1371/journal.pone.0060676

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer (Version 6.0.40). Retrieved from http://www.praat.org/

Bowles, A. R., Chang, C. B., & Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Language Learning*, *66*(4), 774–808. doi:10.1111/lang.12159

Braun, B., & Johnson, E. K. (2011). Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics*, *39*(4), 585–594. doi:10.1016/j.wocn.2011.06.002

Brunelle, M. (2009). Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics*, *37*(1), 79–96. doi:10.1016/j.wocn.2008.09.003

Brunelle, M., Nguyên, D. D., & Nguyên, K. H. (2010). A laryngographic and laryngoscopic study of Northern Vietnamese tones. *Phonetica*, *67*(3), 147–169. doi:10.1159/000321053

Burnham, D., Brooker, R., & Reid, A. (2015). The effects of absolute pitch ability and musical training on lexical tone perception. *Psychology of Music*, *43*(6), 881–897. doi:10.1177/0305735614546359

Burnham, D., Kasisopa, B., Reid, A., Luksaneeyanawin, S., Lacerda, F., Attina, V., … Webster, D. (2015). Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholinguistics*, *36*(6), 1459–1491. doi:10.1017/S0142716414000496

Burnham, D., & Mattock, K. (2007). The perception of tones and phones. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 259–280). Amsterdam: John Benjamins.

Caldwell-Harris, C. L., Lancaster, A., Ladd, D. R., Dediu, D., & Christiansen, M. H. (2015). Factors influencing sensitivity to lexical tone in an artificial language: Implications for second language learning. *Studies in Second Language Acquisition*, *37*(02), 335–357. doi:10.1017/S0272263114000849

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009a). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language*, *108*(1), 1–9. doi:10.1016/j.bandl.2008.02.001

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009b). Sensory processing of linguistic pitch as reflected by the mismatch negativity. *Ear and Hearing*, *30*(5), 552–558. doi:10.1097/AUD.0b013e3181a7e1c2

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, *128*(1), 456–465. doi:10.1121/1.3445785

Chandrasekaran, B., Yi, H.-G., Smayda, K. E., & Maddox, W. T. (2016). Effect of explicit dimensional instruction on speech category learning. *Attention, Perception, & Psychophysics*, *78*(2), 566–582. doi:10.3758/s13414-015-0999-x

Chang, D., Hedberg, N., & Wang, Y. (2016). Effects of musical and linguistic experience on categorization of lexical and melodic tones. *The Journal of the Acoustical Society of America*, *139*(5), 2432–2447. doi:10.1121/1.4947497

Chang, Y. S., Yao, Y., & Huang, B. H. (2017). Effects of linguistic experience on the perception of high-variability non-native tones. *The Journal of the Acoustical Society of America*, *141*(2), EL120–EL126. doi:10.1121/1.4976037

Chao, Y. R. (1930). A system of tone letters. *Le Maître Phonétique*, *45*, 24–27.

Chen, C.-C. (2011). *Monographs on linguistics analysis: A synchronic phonology of Mandarin Chinese*. Retrieved from http://www.jstor.org/stable/23749865

Chen, M. Y. (2000). *Tone Sandhi: Patterns across Chinese dialects*. Cambridge: Cambridge University Press.

Chen, X., Anderson, R. C., Li, W., Hao, M., Wu, X., & Shu, H. (2004). Phonological awareness of bilingual and monolingual Chinese children. *Journal of Educational Psychology*, *96*(1), 142–151. doi:10.1037/0022-0663.96.1.142

Cheung, Y. M. (2011). *Vowels and tones in Meixian Hakka: An acoustic and perceptual study* (PhD dissertation, City University of Hong Kong). Retrieved from http://dspace.cityu.edu.hk/bitstream/2031/6509/1/abstract.html

Chiao, W.-H., Kabak, B., & Braun, B. (2011). When more is less: Non-native perception of level tone contrasts. *Psycholinguistic Representation of Tone Conference*, 42–45. Hong Kong.

Clahsen, H., & Muysken, P. (1986). The availability of universal grammar to adult and child learners - a study of the acquisition of German word order. *Second Language Research*, *2*(2), 93–119. doi:10.1177/026765838600200201

Cooper, A., & Wang, Y. (2010). Cantonese tone word learning by tone and non-tone language speakers. *Interspeech 2010*, 1840–1843. Retrieved from https://pdfs.semanticscholar.org/7ecd/78a05a48922f4f4690ce0888f33073d465d0.pdf

Cooper, A., & Wang, Y. (2011). The influence of tonal awareness and musical experience on tone word learning. *Proceedings of the 17th International Congress of Phonetic Sciences*, 512–515. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Cooper,%20Angela/Cooper,%20Angela.pdf

Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *The Journal of the Acoustical Society of America*, *131*(6), 4756–4769. doi:10.1121/1.4714355

Cooper, A., & Wang, Y. (2013). Effects of tone training on Cantonese tone-word learning. *The Journal of the Acoustical Society of America*, *134*(2), EL133–EL139. doi:10.1121/1.4812435

Cruttenden, A. (1997). *Intonation*. Cambridge: Cambridge University Press.

Curtin, S., & Werker, J. F. (2018). PRIMIR on Tone. *Frontiers in Psychology*, *9*, 1–2. doi:10.3389/fpsyg.2018.01007

Delogu, F., Lampis, G., & Belardinelli, M. O. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *European Journal of Cognitive Psychology*, *22*(1), 46–61. doi:10.1080/09541440802708136

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). Timonium, MD: York Press.

Flege, J. E. (2003). A method for assessing the perception of vowels in a second language. In E. Fava & A. Mioni (Eds.), *Issues in clinical linguistics* (pp. 19–44). Padova: Unipress.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437–470. doi:10.1006/jpho.1997.0052

Fletcher, J., Stirling, L., Mushin, I., & Wales, R. (2002). Intonational rises and dialog acts in the Australian English map task. *Language and Speech*, *45*(3), 229–253. doi:10.1177/00238309020450030201

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*(2), 268–294. doi:10.1016/j.wocn.2007.06.005

Fu, Q.-J., & Galvin, J. J. (2007). Perceptual learning and auditory training in cochlear implant recipients. *Trends in Amplification*, *11*(3), 193–205. doi:10.1177/1084713807301379

Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, *21*(1), 1–33. doi:10.1177/002383097802100101

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, G. D. (2000). A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience*, *12*(1), 207–222. doi:10.1162/089892900561841

Gfeller, K., Guthe, E., Driscoll, V., & Brown, C. J. (2015). A preliminary report of music-based training for adult cochlear implant users: Rationales and development. *Cochlear Implants International*, *16*(sup3), S22–S31. doi:10.1179/1467010015Z.000000000269

Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, *25*(2), 207–231. doi:10.1006/jpho.1997.0042

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. doi:10.1177/00238309030460040201

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America*, *107*(5), 2711–2724. doi:10.1121/1.428657

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (pp. 57–77). Amsterdam: John Benjamins Publishing.

Guy, G., Horvath, B., Vonwiller, J., Daisley, E., & Rogers, I. (1986). An intonational change in progress in Australian English. *Language in Society*, *15*(1), 23–51.

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421. doi:10.1016/S0095-4470(03)00016-0

Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*(2), 269–279. doi:10.1016/j.wocn.2011.11.001

Hashimoto, M. J. (2010). *The Hakka dialect: A linguistic study of its phonology, syntax and lexicon*. Cambridge: Cambridge University Press.

Hutka, S., Bidelman, G. M., & Moreno, S. (2015). Pitch expertise is not created equal: Cross-domain effects of musicianship and tone language experience on neural and behavioural discrimination of speech and music. *Neuropsychologia*, *71*, 52–63. doi:10.1016/j.neuropsychologia.2015.03.019

Hwa-Froelich, D., Hodson, B. W., & Edwards, H. T. (2002). Characteristics of Vietnamese phonology. *American Journal of Speech-Language Pathology*, *11*, 264–273. doi:doi:10.1044/1058-0360(2002/031)

Ingvalson, E. M., Barr, A. M., & Wong, P. C. M. (2013). Poorer phonetic perceivers show greater benefit in phonetic-phonological speech learning. *Journal of Speech, Language and Hearing Research*, *56*(3), 1045–1050. doi:10.1044/1092-4388(2012/12-0024)

Jones, J. L., Lucker, J., Zalewski, C., Brewer, C., & Drayna, D. (2009). Phonological processing in adults with deficits in musical pitch recognition. *Journal of Communication Disorders*, *42*(3), 226–234. doi:10.1016/j.jcomdis.2009.01.001

Kaan, E., Barkley, C. M., Bao, M., & Wayland, R. (2008). Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: An event-related potentials training study. *BMC Neuroscience*, *9*, 53. doi:10.1186/1471-2202-9-53

Kaan, E., Wayland, R., Bao, M., & Barkley, C. M. (2007). Effects of native language and training on lexical tone perception: An event-related potential study. *Brain Research*, *1148*, 113–122. doi:10.1016/j.brainres.2007.02.019

Kaan, E., Wayland, R., & Keil, A. (2013). Changes in oscillatory brain networks after lexical tone training. *Brain Sciences*, *3*(2), 757–780. doi:10.3390/brainsci3020757

Kirby, J. (2010). Dialect experience in Vietnamese tone perception. *The Journal of the Acoustical Society of America*, *127*(6), 3749–3757. doi:10.1121/1.3327793

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010a). Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity. *Brain Research*, *1313*, 124–133. doi:10.1016/j.brainres.2009.11.061

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010b). The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics*, *23*(1), 81–95. doi:10.1016/j.jneuroling.2009.09.001

Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. A. (2004). Human frequency-following response: Representation of pitch contours in Chinese tones. *Hearing Research*, *189*(1), 1–12. doi:10.1016/S0378-5955(03)00402-7

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2). doi:10.1111/j.1467-7687.2006.00468.x

Ladd, D. R. (2008). *Intonational Phonology*. Cambridge: Cambridge University Press.

Lee, C.-Y., & Hung, T.-H. (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, C.-Y., Tao, L., & Bond, Z. S. (2010). Identification of multi-speaker Mandarin tones in noise by native and non-native listeners. *Speech Communication*, *52*(11), 900–910. doi:10.1016/j.specom.2010.01.004

Lee, W.-S., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association*, *33*(1), 109–112. doi:10.1017/S0025100303001208

Lee, W.-S., & Zee, E. (2009). Hakka Chinese. *Journal of the International Phonetic Association*, *39*(01), 107. doi:10.1017/S0025100308003599

Liu, F., Jiang, C., Francart, T., Chan, A. H. D., & Wong, P. C. M. (2017). Perceptual learning of pitch direction in congenital amusia. *Music Perception*, *34*(3), 335–351. doi:10.1525/mp.2017.34.3.335

Liu, F., Jiang, C., Thompson, W. F., Xu, Y., Yang, Y., & Stewart, L. (2012). The mechanism of speech processing in congenital amusia: Evidence from Mandarin speakers. *PloS One*, *7*(2), e30374. doi:10.1371/journal.pone.0030374

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: Discrimination, identification and imitation. *Brain*, *133*(6), 1682–1693. doi:10.1093/brain/awq089

Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, *133*(2), 385–394. doi:10.1016/j.cognition.2014.06.004

Liu, L., & Kager, R. (2018). Monolingual and bilingual infants' ability to use non-native tone for word learning deteriorates by the second year after birth. *Frontiers in Psychology*, *9*(117). doi:10.3389/fpsyg.2018.00117

Lo, C. Y., McMahon, C. M., Looi, V., & Thompson, W. F. (2015). Melodic contour training and its effect on speech in noise, consonant discrimination, and prosody perception for cochlear implant recipients. *Behavioural Neurology*, *2015:352869*. doi:10.1155/2015/352869

Lu, S., Wayland, R., & Kaan, E. (2015). Effects of production training and perception training on lexical tone perception – A behavioral and ERP study. *Brain Research*, *1624*, 28–44. doi:10.1016/j.brainres.2015.07.014

Maddieson, I., Bhattacharya, T., Smith, D. E., & Croft, W. (2011). Geographical distribution of phonological complexity. *Linguistic Typology*, *15*(2), 267–279. doi:10.1515/lity.2011.020

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. doi:10.1044/1092-4388(2007/067)

Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language: Association with lexical tone agnosia. *Brain*, *133*(9), 2635–2642. doi:10.1093/brain/awq178

Nguyen, S., Tillmann, B., Gosselin, N., & Peretz, I. (2009). Tonal language processing in congenital amusia. *Annals of the New York Academy of Sciences*, *1169*(1), 490–493. doi:10.1111/j.1749-6632.2009.04855.x

Nguyễn, V. L., & Edmondson, J. A. (1997). Tones and voice quality in modern northern Vietnamese: Instrumental case studies. *Mon-Khmer Studies, 28,* 1–18. doi:10.15144/MKSJ-28.1

Norman, J. (1988). *Chinese*. Cambridge: Cambridge University Press.

Paribakht, T. S., & Wesche, M. (1996). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition* (pp. 174–200). doi:10.1017/CBO9781139524643.013

Peretz, I. (2001). Brain specialization for music. *Annals of the New York Academy of Sciences*, *930*(1), 153–165. doi:10.1111/j.1749-6632.2001.tb05731.x

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472. doi:10.1121/1.3593366

Pierrehumbert, J., & Hirschberg, J. B. (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). Retrieved from https://academiccommons.columbia.edu/doi/10.7916/D8KD24FP

Plack, C. J., Oxenham, A. J., Fay, R. R., & Popper, A. N. (Eds.). (2006). *Pitch: Neural Coding and Perception*. New York: Springer.

Qin, Z., & Jongman, A. (2016). Does second language experience modulate perception of tones in a third language? *Language and Speech*, *59*(3), 318–338. doi:10.1177/0023830915590191

Qin, Z., & Mok, P. P. K. (2013). Discrimination of Cantonese tones by speakers of tone and non-tone languages. *Kansas Working Papers in Linguistics*, *34*. doi:10.17161/KWPL.1808.12864

Quam, C., & Creel, S. C. (2017). Tone attrition in Mandarin speakers of varying English proficiency. *Journal of Speech, Language and Hearing Research*, *60*(2), 293–305. doi:10.1044/2016_JSLHR-S-15-0248

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *The Journal of the Acoustical Society of America*, *105*(1), 512–521. doi:10.1121/1.424522

Reid, A., Burnham, D., Kasisopa, B., Reilly, R., Attina, V., Rattanasone, N. X., & Best, C. T. (2015). Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Attention, Perception, & Psychophysics*, *77*(2), 571–591. doi:10.3758/s13414-014-0791-3

Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology*, *5*(1318). doi:10.3389/fpsyg.2014.01318

Saengtummachai, W. (2003). *A phonological study of the Meixian Hakka dialect in Bangkok, Thailand, in comparison with Hashimoto's study of the Meixian Dialect in China* (Masters thesis). Mahidol University, Thailand.

Schaefer, V., & Darcy, I. (2014). Lexical function of pitch in the first language shapes cross-linguistic perception of Thai tones. *Laboratory Phonology*, *5*(4), 489–522. doi:10.1515/lp-2014-0016

Shih, C. (1988). Tone and intonation in Mandarin. *Work Papers of the Cornell Phonetic Laboratory*, *3*, 83–109.

Showalter, C. E., & Hayes-Harb, R. (2013). Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Second Language Research*, *29*(2), 185–200. doi:10.1177/0267658313480154

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273–293. doi:10.1177/0023830909357156

So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, *36*(2), 195–221. doi:10.1017/S0272263114000047

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. doi:10.1162/jocn.2008.20131

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon/Pearson Education.

Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., & Peretz, I. (2011). Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages. *Frontiers in Psychology*, *2*(120), 1–15. doi:10.3389/fpsyg.2011.00120

Tong, X., & Tang, Y. C. (2016, May 31). *Modulation of musical experience and prosodic complexity on lexical pitch learning*. Presented at Speech Prosody, 217–221. doi:10.21437/SpeechProsody.2016-45

Tsukada, K., Xu, H. L., & Rattanasone, N. X. (2015). The perception of Mandarin lexical tones by listeners from different linguistic backgrounds. *Chinese as a Second Language Research*, *4*(2), 141–161. doi:10.1515/caslar-2015-0009

Vũ, T. P. (1982). Phonetic properties of Vietnamese tones across dialects. In D. Bradley (Ed.), *Papers in Southeast Asian Linguistics* (Vol. 8, pp. 55–75). Sydney: Australian National University.

Wang, X. (2013). Perception of Mandarin tones: The effect of L1 background and training. *The Modern Language Journal*, *97*(1), 144–160. doi:10.1111/j.1540-4781.2013.01386.x

Wang, Y., Jongman, A., & Sereno, J. A. (2001). Dichotic perception of Mandarin tones by Chinese and American listeners. *Brain and Language*, *78*(3), 332–348. doi:10.1006/brln.2001.2474

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*(2), 1033–1043. doi:10.1121/1.1531176

Wang, Y., Sereno, J. A., Jongman, A., & Hirsch, J. (2003). fMRI evidence for cortical modification during learning of Mandarin lexical tone. *Journal of Cognitive Neuroscience*, *15*(7), 1019–1027. doi:10.1162/089892903770007407

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. doi:10.1121/1.428217

Wayland, R., & Guion, S. (2003). Perceptual discrimination of Thai tones by naive and experienced learners of Thai. *Applied Psycholinguistics*, *24*(1), 113–129. doi:10.1017/S0142716403000067

Wayland, R., & Guion, S. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*(4), 681–712. doi:10.1111/j.1467-9922.2004.00283.x

Wayland, R., Herrera, E., & Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics*, *38*(4), 654–662. doi:10.1016/j.wocn.2010.10.001

Wayland, R., & Li, B. (2005). *Training native Chinese and native English listeners to perceive Thai tones*. Presented at the ISCA workshop on plasticity in speech perception (PSP 2005). Retrieved from http://www.isca-speech.org/archive_open/psp_2005/psp5_062.html

Wayland, R., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, *36*(2), 250–267. doi:10.1016/j.wocn.2007.06.004

Wayland, R., Zhu, Y., & Kaan, E. (2015, August 10). *Perception of pitch contours by native and non-native tone listeners*. Presented at the 18th International Congress of Phonetic Sciences, Glasgow, UK. Retrieved from http://www.icphs2015.info/pdfs/Papers/ICPHS0493.pdf

Wei, C.-G., Cao, K., & Zeng, F.-G. (2004). Mandarin tone recognition in cochlear-implant subjects. *Hearing Research*, *197*(1–2), 87–95. doi:10.1016/j.heares.2004.06.002

Werker, J. F., & Curtin, S. (2005). PRIMIR: A Developmental Framework of Infant Speech Processing. *Language Learning and Development*, *1*(2), 197–234. doi:10.1080/15475441.2005.9684216

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(4), 565–585. doi:10.1017/S0142716407070312

Wong, P. C. M., 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. doi:10.1038/nn1872

Wu, X., & Lin, H. (2008). Perception of Mandarin tones by Mandarin and English listeners. *Journal of Chinese Language and Computing*, *18*(4), 175–187.

Wu, X., Munro, M. J., & Wang, Y. (2014). Tone assimilation by Mandarin and Thai listeners with and without L2 experience. *Journal of Phonetics*, *46*, 86–100. doi:10.1016/j.wocn.2014.06.005

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, *120*(2), 1063–1074. doi:10.1121/1.2213572

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

Zhang, J. (2004). The role of contrast-specific and language-specific phonetics in contour tone distribution. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically Based Phonology* (pp. 157–191). Cambridge: Cambridge University Press.

Zhao, T. C., & Kuhl, P. K. (2015). Higher-level linguistic categories dominate lower-level acoustics in lexical tone processing. *The Journal of the Acoustical Society of America*, *138*(2), EL133–EL137. doi:10.1121/1.4927632

Zheng, H.-Y., Minett, J. W., Peng, G., & Wang, W. S.-Y. (2012). The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study. *Language and Cognitive Processes*, *27*(2), 184–209. doi:10.1080/01690965.2010.520493