# Cross-language Speech Perception and Foreign Language Acquisition: Adult Perception of Thai and Mandarin Tonal Contrasts by Naïve and Mandarin-learning American English Speakers

Samuel J. Alfieri

New York University

**Abstract**

This study investigates the effects of adult second language (L2) acquisition on perception of non-native phonological categories. Specifically, monolingual American English speakers, and American English speakers who have learned Mandarin in adulthood were tested on perceptual discrimination of Thai and Mandarin tonal contrasts. This research question extends previous work by Wayland and Guion (2004) as well as Schaefer and Darcy (2013), who found that native Mandarin speakers perceive Thai tones significantly more accurately than native American English speakers. However, it remains an open question as to whether L2 acquisition of Mandarin would confer a congruent perceptual advantage on learners. If Mandarin learners form new phonological categories for tone, they should perceive both Mandarin and Thai tonal contrasts better than American English speakers who have no experience in a language with lexical tone. The findings indicate no significant difference in accuracy rates between naïve listeners and Mandarin learners for Mandarin or Thai tonal contrasts, suggesting that even if phonological categories for tone can be established, it is difficult to generalize them to a third language. However, differences in accuracy rates between individual contrasts are significant, and are interpreted in light of the Speech Learning Model (Flege, 1995), and the Perceptual Assimilation Model (Best, 1995). Strong performance on tone contrasts that differ markedly in contour shape and direction over level tone contrasts supports previous researchers' claims that intrinsic differences between level and contour tones affect nonnative perception (Francis, Ciocca, & Ng, 2003; Lee, Vakoch, and Wurm, 1996; Qin & Mok, 2012; Wayland & Guion, 2004).

## 1. Introduction

1.1 Review of literature in L2 speech perception

For an adult to produce and perceive nonnative segmental speech sounds presents difficulties that are not found when performing the same tasks with native speech sounds. Studies investigating production and perception of non-native consonant and vowel (segmental) contrasts have found that some are quite difficult to discriminate (i.e., /ʏ/ vs. /ʊ/) whereas others are easily discriminable (i.e. lateral vs. apical clicks) despite listeners’ lack of familiarity with the sounds (Best, McRoberts, & Sithole, 1988; Polka, 1995; Werker & Tees, 1984). First language (L1) interference in acquisition of an L2 usually manifests itself transparently as foreign-accented speech in adult learners, with difficulties arising when L2 contrasts are not salient in the learner’s L1 (Flege, Schirru, & MacKay, 2003; Flege & Davidian, 1984; Flege, McCutcheon, & Smith, 1987; Guion, Flege, Akahane-Yamada, & Pruitt, 2000; Jamieson & Morosan, 1986; Logan, Lively, & Pisoni, 1991; Polka, 1995).

The critical period hypothesis predicts that the plasticity of the human language faculty deteriorates significantly with age. However, studies show that with simple training, even adult speakers can acquire nonnative phonemic contrasts (Jamieson & Morosan, 1986; Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan et al., 1991; Pisoni, Aslin, Perey, & Hennessy, 1982; Wayland & Guion, 2004). McClaskey, Pisoni, and Carrell (1983) furthered this work and found that American English speakers who were trained to discriminate a three-way contrast in voice onset time between voiceless aspirated, voiceless unaspirated, and voiced stops at one place of articulation (POA) were able to transfer their perceptual learning to another POA without any additional training. The present study asks if

American English speakers who have acquired Mandarin tone as adults are able to transfer their perceptual learning to novel phonological contrasts in Thai.

Alternatively, salient linguistic cues may be untransferrable from Mandarin to Thai for nontone language speakers. In the previous example, transfer of the perceptual distinction is based on ability to distinguish VOT in consonant-vowel (CV) syllables. Studying the suprasegmental feature of tone may not yield the same result.

1.2 The tone systems of Thai and Mandarin

The Mandarin tone system consists of four phonemic tones: tone 1 (T1) is a high-level tone, tone 2 (T2) is a high-rising tone, tone 3 (T3) is a low-falling tone, and tone 4 (T4) is a high-falling tone. While Mandarin has only one level tone, Thai has three. The Thai tonal system consists of five tones: a rising tone (R), falling tone (F), low-level tone (L), mid-level tone (M), and high-level tone (H) (Wayland & Guion, 2004).

Burnham, Kirkwood, Luksaneeyanawin, and Pansottee, (1992) suggest that initial pitch level is the most salient linguistic cue for Australian English and Thai speaking individuals in distinguishing Thai tones, however a variety of studies on Mandarin suggest that various other linguistic cues such as F0 height, contour, turning point, and vowel duration are most salient when discriminating different tone contrasts (Gandour, 1984; Gårding, Kratochvil, Svantesson, & Zhang, 1986; Massaro & Cohen, 1985; Moore & Jongman, 1997; Shen & Lin, 1991). In perceiving nonnative segmental contrasts, listeners do not always attend to acoustic cues that are relevant in the target language, a finding that may hold true for suprasegmental contrasts as well (Iverson et al., 2003). Although native Mandarin speakers demonstrate better perceptual discrimination of Thai tonal contrasts than American English speakers, nonnative acquisition of Mandarin tone may not be robust enough to confer a significant advantage in perceiving Thai contrasts if listeners do not generalize their learning and attend to relevant features like native tone language speakers (Schaefer & Darcy, submitted 2013; Wayland & Guion, 2004). Differences in salient cues when discriminating Thai and Mandarin tonal contrasts may require that nonnative learners acquire each tone system individually to achieve robust discriminatory abilities.

1.3 Theories of categorical vs. non-categorical perception of tone

Some studies involving manipulated stimuli on continua between different tones show that tone language speakers have formed sharper categorical boundaries and perceive differences between similar stimuli more accurately than do nontone language speakers (Hallé, Chang, & Best, 2004; Xu, Gandour, & Francis, 2006), although Wu and Lin (2008) found Canadian English speakers to outperform native Mandarin speakers in a similar task. It has been suggested that perception of lexical tone may be more similar to that of the vowel space than consonantal contrasts (Abramson, 1976). Iverson and Kuhl (2000) tested native American English speakers on an /i/-/e/ continuum and found a perceptual magnet effect to occur, but observed no sharp phoneme boundaries between categories. This contrasts with Wood (1976), who observed sharp increases in discriminability near phoneme boundaries for VOT-manipulated stop consonants. Furthermore, perception of level tones seems to be primarily noncategorical, whereas dynamic tone contours may be perceived categorically (Abramson, 1977; Francis et al., 2003; W. S. Wang, 1976).

Hallé, Chang, and Best (2002) tested Taiwanese and French listeners on Mandarin Taiwanese tonal contrasts along continua. Patterns of response suggest that for French listeners who are naïve to lexical tone, perception of tonal contours is psychophysically based, whereas for Mandarin Taiwanese listeners, whose tone system very closely resembles that of Mandarin

Chinese, perception is heavily influenced by native phonology and tone is perceived categorically.

Further evidence suggests that the mode by which tone is processed may be dependant on the specific acoustic features of the level tone or tone contour in question. Early research by Wang (1976) and Abramson (1977) presented seemingly conflicting theories, with Wang finding that Mandarin tone was discriminated categorically on a continuum whereas Abramson found that Thai was perceived noncategorically using similar methods. Francis et al. (2003) sought to resolve this debate by testing native speakers of Cantonese on a variety of contrasts that resembled both level Thai tones as well as contoured Mandarin tones. Their findings replicated Wang (1976) and Abramson (1977), with level Cantonese tones being perceived non-categorically in contrast to contoured Cantonese tones, which were perceived categorically. This reaffirms scholars' suggestions that perception of level tones is systematically more difficult than dynamic tones (Abramson, 1977; Hallé et al., 2004; L. Lee & Nusbaum, 1993). The Cantonese contrast of low-rising vs. high-rising has been noted as an exception to this generalization, wherein both non-native and native speakers exhibit significant difficulty perceiving a categorical difference between the two (Francis et al., 2003). The lack of a distinct difference in the overall shape of the tone contours for this distinction, like level tones, makes it intrinsically more difficult than tone contours that vary drastically in shape.

The lack of categoricity in perceiving tonal contrasts that non-tone language speakers demonstrate seems to be due to their lack of exposure to lexical tone that marks a phonological difference between segments. Fundamental differences in the way that nontone vs. tone L1 speakers perceive tones, as well as in the way that different types of tonal contrasts are perceived, could complicate any effort to explain cross-linguistic tone perception using a single model.

1.4 Cross-linguistic perception of tone

Although native Mandarin speakers perceive Thai tone better than American English speakers, this does not mean that cross-linguistically, speakers of tone languages always perform better at lexical tone discrimination tasks than nontone language speakers. Another study by Lee, Vakoch, and Wurm (1996) found that while Cantonese listeners perceived Mandarin tone better than English speakers in a same-different AX task, Mandarin speakers were no better at distinguishing Cantonese tones than English speakers. Qin and Mok (2012) found Mandarin speakers to exhibit more difficulty in perceiving the Cantonese T3-T6 (mid level vs. low level) contrast than both English and French speakers, suggesting that L1 Mandarin experience in a language with only one level tone confers a perceptual disadvantage in perceiving nonnative level tonal contrasts versus non-tone language speakers. This hypothesis is supported by the findings of Chiao, Kabak, and Braun (2011), wherein tone language speakers possessing two level tones in their native phonology conferred more difficulties in perceiving a four-way level tonal contrast than tone language speakers possessing one level tone. Both groups performed poorer than nontone language speakers, who did not assimilate lexical tone to native categories.

So (2006) tested Cantonese (a tone language) and Japanese (a pitch accent language) speakers on Mandarin tonal contrasts, and found that her participants had similar difficulties with two of the same contrasts (T2-T3 and T1-T4) as non-tone language speakers. Cantonese listeners made more errors than the Japanese listeners in both of these conditions, although the difference between the groups was only strongly significant for the T1-T4 condition. So explains this result as a case of single category assimilation using Best’s Perceptual Assimilation Model (PAM), which accounts for native language interference in perception of nonnative speech sounds and posits that Cantonese speakers assimilated both Mandarin T1 and T4 to their native Cantonese T1 category (Best, 1994; Best, 1995; Best, McRoberts, & Goodell, 2001; Best & Tyler, 2007).

### 1.4.1 Best's Perceptual Assimilation Model (PAM)

The PAM predicts that non-native speech perception is strongly affected by the listener's native language, and that accuracy in a perceptual task is influenced by perceptual assimilation of non-native sounds to native categories. The PAM posits that the relative accuracy of listeners' responses is predictable based on the type of assimilation that non-native sounds undergo for the listener. There are five categorical assimilations patterns that may occur: single category, two category, category goodness, uncategorized-categorized, and uncategorized-uncategorized. In *single category (SC) assimilation*, two non-native sound categories are mapped onto a single native category, and discrimination is predicted to be very poor, as no phonemic difference between the stimuli is perceived. This contrasts with *two category (TC) assimilation*, in which two non-native sound categories are mapped onto two different native categories, and discrimination is expected to be excellent, congruent to accuracy in distinguishing the native categories. Alternatively, two sounds may assimilate to the same native category, however one assimilates significantly better than the other, which is perceived as a more deviant exemplar. This constitutes a *category goodness (CG) difference in assimilation*, with discrimination anticipated to be moderate to excellent. Some nonnative speech sounds may not assimilate to native categories, and result in a *uncategorized-categorized (UC) assimilation*, with discrimination expected to be good to excellent. An additional possibility is that both nonnative speech sounds are uncategorized, in an *uncategorized-uncategorized (UU) assimilation*. Best and Tyler (2007) predict discrimination for UU assimilation contrasts to range from poor to moderate depending on their acoustic characteristics. Finally, some speech sounds may be *non-assimilable* and be perceived as nonspeech sounds, such as Zulu clicks (Best et al., 1988). Discrimination on contrasts involving non-assimilable sounds is expected to be good to excellent, again depending on their specific acoustic properties. Guion et al. (2000) have additionally proposed a revision to the PAM to account for UC assimilation where the uncategorized sound is close in phonological space to the categorized sound. They would predict discrimination to be relatively poorer than the good to excellent level that the original model proposes.

This model of PAM, however, fails to account for how suprasegmental features of speech, such as tone, are assimilated into native speech categories and perceived by listeners. Building on Best's original framework, Hallé, Chang, and Best (2004) suggest that lexical tones may be perceived either as uncategorized speech or as nonspeech. Thus, testing English speakers on lexical contrasts with words that contain the same segments, differing only in suprasegmental tone, would constitute cases of UU assimilation, and discrimination would be predicted to range from fair to good, depending on the perceived salience of the phonetic differences. This prediction is unlikely in the context of the current study, since participants were instructed that they would be participating in a linguistics experiment and listening to speech sounds.

Additionally, previous studies have found that even for non-tone language speakers, speech sounds with lexical tone are processed linguistically rather than psychoacousticaly. Discrimination on speech sounds is poorer for non-tone language speakers than on low-pass filtered speech stimuli or musical stimuli matched in pitch to lexical tone contrasts (Burnham et al., 1996). Further evidence for linguistic perception of tone comes from Van Lancker and Fromkin (1973), who find that Thai speakers demonstrate an advantage in discriminating tones presented to the right ear (processed in the left hemisphere of the brain, which dominates language processing) versus the left. Presentation of hummed speech did not produce the same linguistic bias in Thai speakers, however nor did non-tone English speakers confer a similar advantage when speech sounds were presented to the right ear. In the context of the present study, this could mean that participants will not attune to the suprasegmental features of lexical tone linguistically. This is not in contrast to Hallé, Chang, and Best (2002), who did not address the question of linguistic vs. non-linguistic perception of tone. The best explanation of the difference in modes of processing comes from Werker and Tees (1984), who posit that there is an intermediate level of processing between the phonological and acoustic levels, the phonetic level. Whereas the Taiwanese listeners in Hallé, Chang, and Best (2002) categorically assimilated Mandarin tone to native phonological categories, French listeners failed to attach salience to the speaker's pitch and responded according to psychoacoustic phonetic cues.

Phonetic processing could account for good discrimination of nonnative sounds when it prevails over phonemic processing, however psychoacoustic cues that distinguish difficult tonal contrasts are less accessible to nontone language speakers when linguistic processing takes place (Burnham et al., 1996). Linguistic processing in Mandarin learners could thus result in decreased perceptual discrimination of tonal contrasts.

It is not the case, however, that intonational contrasts have no salience in English. As a stress-accent language, segmentally identical utterances spoken with differing intonation may contrast in meaning (Beckman & Pierrehumbert, 1986). For example, a down-step in intonational pitch on the sentential level usually indicates a declarative sentence in American English, whereas rising intonation indicates that a sentence is a question. Since intonation is salient on the sentential level, it is possible that acoustic cues for lexical tone in Mandarin and

Thai would be categorized according to salient intonation patterns for American English listeners, or *i-Categories* (So & Best, 2011).

1.4.2 Flege's Speech Learning Model (SLM)

Another primary question of the current investigation is whether Mandarin learners are able to generalize their acquisition of lexical tone to unfamiliar Thai stimuli. I turn to an alternative model of language acquisition, Flege's Speech Learning Model (SLM) to propose a possible mechanism for this to occur. The SLM predicts that difficulties in producing and perceiving nonnative phoneme categories arise from age-related changes in how L1 and L2 phonetic subsystems interact (Flege, 1995; Flege, Schirru, & MacKay, 2003; Flege & Davidian, 1984; Flege, McCutcheon, & Smith, 1987). The SLM predicts that as L1 vowel and consonant categories develop and become robust, acoustically similar L2 sounds will assimilate to L1 categories. By adulthood, L1 categories have solidified to the extent that beginning adult L2 learners tend to perceive and produce L2 using the categorically assimilated L1 segments rather than their original phonetic realizations. Like PAM, SLM predicts that non-assimilable sounds are easy to discriminate, since they are not perceptually shaped by assimilation to L1 categories.

The SLM, unlike the PAM, also accounts for the formation of new categories throughout life (Flege et al., 2003). The difficulty in innovating new categories in L2 comes from L2 sounds' assimilation onto L1 categories, so the more dissimilar an L2 speech sound is from its closest L1 speech sound, the more likely it will be that a new category will develop. If two sounds are too close in perceived similarity, assimilation of the L2 sound to the L1 category will block the formation of a new, robust category, and a "merged" category may develop over time that subsumes both the L1 and L2 speech sounds. It is possible that segments containing lexical tone in Thai or Mandarin are too similar to American English L1 categories to allow for the robust formation of distinct tone categories in Mandarin learners.

## 1.5 Predictions for the current study

For native American English-speaking learners of Mandarin, are tone categories established? Mandarin learners are explicitly taught that tone is a salient contrast, so learners have been specifically instructed to attend to suprasegmental differences. Applying the SLM and PAM to tone acquisition, one might predict tones that resemble American English sentential intonation contours to be assimilated to L1 intonational categories. However, neither model explicitly makes predictions about the conflation of tone and intonation, and the relationship between the two is still unclear. If we predict that they may be conflated, following the findings of So and Best (2011), Mandarin T2 and T3 and Thai rising tones might be assimilated to the American English intonation pattern that indicates questioning, and Mandarin T4 and Thai falling tone might be assimilated to the declaratory American English intonational pattern, predicting the best discrimination for T2-T4, T3-T4, and RF contrasts, all examples of TC assimilation. The SLM would predict, however, that non-native and non-assimilable lexical tones would constitute novel categories, and discrimination would be good even with little training independent of the types of assimilation proposed by the PAM.

The SLM also predicts that practice and experience with L2 categories will result in robust category formation. If this is the case, combining SLM with PAM, we would expect to find that experienced Mandarin learners have formed tonal categories and behave similarly to native Mandarin speakers. If category formation has taken place, we would expect Mandarin learners to possess a greater ability to discriminate Thai tones than their monolingual American English-speaking peers, as tonal contrasts are assimilated in TC, CG, or UC patterns, depending on the contrast in question. All of these assimilation patterns predict discrimination to be at least moderate, in comparison to the fair to good discrimination predicted by UU assimilation. Alternatively, robust category formation may not take place, even if a perceptual advantage is observed for Mandarin learners. Some auditory training studies have successfully taught American English and other tone language speakers to improve their perception of nonnative tonal contrasts, however their performance still falls below that of native speakers (Y. Wang, Spence, Jongman, & Sereno, 1999; Wayland & Guion, 2004). This suggests that even if learners achieve an intermediate level of acquisition that lies between naïveté and native L1 fluency, a congruent advantage in perceiving a third language's tonal contrasts may not be conferred.

## 2. Methods

### 2.1 Subjects

Participants included 20 native speakers of American English (4 male, 16 female, mean age = 20.3) and 15 American English-speakers who went on to study Mandarin Chinese as adults (10 male, 5 female). Two of the male subjects from the Mandarin learners group were excluded, one because of a computer error, and the other because he demonstrated irregular American English prosody in his speech as a result of input from British English-speaking parents and relatives. The average age of the Mandarin learners included in analysis was 20.9 years, and the mean number of years they had studied Mandarin was 3.9. Although no subjects were native speakers of a language other than American English, those in the Mandarin-naïve group had various levels of exposure to Arabic, Czech, French, German, Hebrew, Hindi, Italian, Polish, Portuguese, and/or Spanish. Three subjects reported fluency in French and one in Spanish. All participants in the second group were second language learners of Mandarin Chinese; while four of them had no significant exposure to other languages, other language exposure was reported for French, Hebrew, Italian, Latin, and Russian. Two Mandarin-learners reported fluency in French. None of the languages that any participants reported learning, other than Mandarin, exhibit phonological tonal contrasts.

All participants were affiliated with New York University at the time of study. Of those included in the analysis, 31 were undergraduate students, one was a graduate student, and one was a university employee.

2.2 Stimuli

The speech materials used in this study were obtained from other researchers and consisted of Thai and Mandarin monosyllables, each spoken by both a male and female speaker. Trials were constructed by playing three monosyllables consecutively with a 750ms interstimulus interval. Triads were presented using an ABX design. The first two sounds were two different monosyllables uttered by a female speaker, followed by a male speaker uttering a monosyllable that matched either the first or second utterance of the female speaker. Test stimuli differed by suprasegmental tonal contrasts, and control stimuli differed in voicing, place, manner, vowel, or coda consonant.

The Thai stimuli were the same as those used by Schaefer and Darcy (submitted, 2013). They were recorded in isolation without a carrier phrase by speakers of the Central Thai dialect (i.e. standard Thai) and all stimuli used in the current study were the second of three repetitions that were recorded. Since Thai does not contain complete sets of real words for the same syllabi with all 5 contrastive tones, some stimuli were non-words, however Schaefer and Darcy's native Thai speakers were still able to discriminate non-word tonal contrasts with near-perfect accuracy. This information taken with the fact that none of the participants in the current study are Thai speakers, and should not have lexical access to Thai words, suggests that non-word stimuli should present no inherent disadvantage to listeners over listening to real word stimuli, and all were pooled in the final analysis.

Mandarin stimuli were obtained from Charles Chang. Syllabi were recorded by Native Mandarin speakers from the northern part of mainland China who were living in the Washington, DC area. They were made using a Zoom H4n recorder and Audix HT5 head-mounted condenser microphone. Stimuli were chosen for clarity from multiple recordings of identical monosyllables.

The Thai block of the experiment consisted of 288 distinct stimuli, separated into 96 ABX trials. Forty-eight of these trials tested participants on their perception of 10 Thai tonal contrasts, and 48 control trials differed by voicing, place, manner, vowel, or coda. The Mandarin block consisted of 144 distinct stimuli separated into 48 ABX trials. Twenty-four trials tested 6 different Mandarin tonal contrasts, and 24 control trials differed by voicing, place, or manner contrasts (see appendix for lists of experimental triads).

2.3 Procedure

To begin, participants were asked to fill out a demographic language questionnaire prior to the experimental task (see appendix). They were then brought to another room to complete the experimental task on a computer.

Participants completed an ABX discrimination task by listening to stimuli through headphones and responding on a button box. They first completed a practice round with feedback using unrelated stimuli to familiarize themselves with the format of the task. They then proceeded to the Thai block of the task. As stimuli were presented, the next trial began as soon as the listener’s response was recorded. After the Thai block, participants were given the option to take a break; when desired, they began the Mandarin block. After they were finished, they were debriefed and compensated for their participation.

Participants began with the Thai task to avoid a training effect bias on the Thai tonal contrasts. Since the experimental group was naïve to Thai but not to Mandarin, they would be expected to exhibit a greater learning effect after being presented with familiar stimuli than the naïve listeners from the control group. Since all participants were naïve to Thai prior to the experiment, any training effect on their Mandarin responses would be expected to remain consistent across conditions.

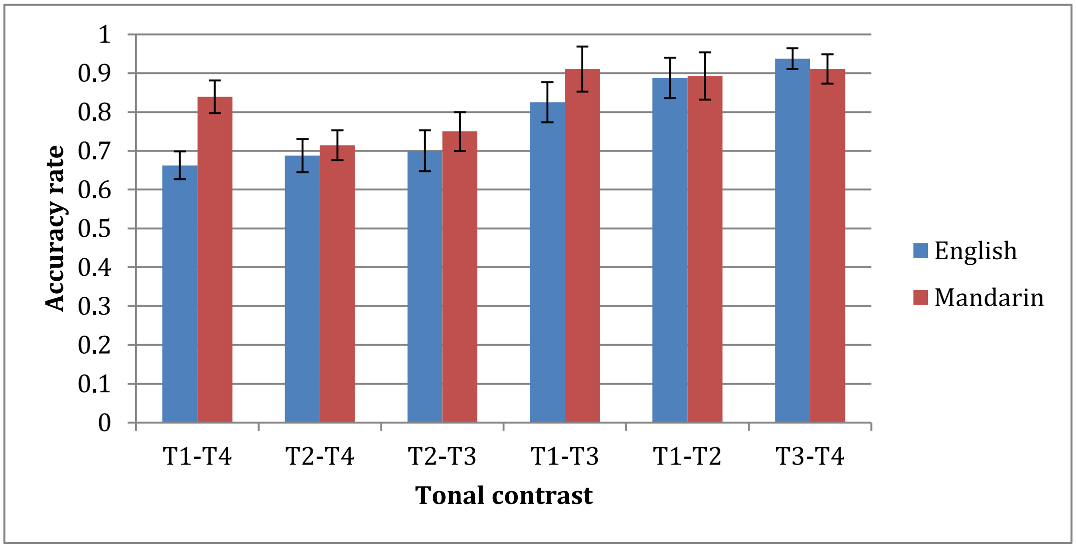
To further control for any directional biases, responses were counterbalanced so that there were an equal number of correct "1" responses as "2." Additionally, stimuli order was randomized for each listener to control for any order effects.

**3. Results**

Analysis of the accuracy data is carried out using a mixed-effects binomial logistic regression implemented in R (Bates, Maechler, & Bolker, 2012).

3.1 Mandarin accuracy

**Figure 1. Accuracy rates for Mandarin by tonal contrast**




In analysis of Mandarin accuracy data, the reference levels are English monolingual for the Language factor and T1-T2 contrast for the Contrast factor. There is no significant effect of language (ß =0.118, z = 0.183, p = .855), nor are there any interactions between tone contrast and language. This means that the Mandarin learners did not perform significantly better or worse than the control naïve English speakers who participated here. However, as shown in table 1, there are significant effects for the contrasts T1-T4 and T2-T3 when T1-T2 is taken as a reference.

**Table 1. Results of mixed-effect binomial regression for Mandarin tonal contrasts**

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 2.16426 | 0.49679 | 4.356 | 0.0000132*** |
| LanguageMan | 0.11834 | 0.64805 | 0.183 | 0.85511 |
| contrast13 | -0.56167 | 0.5038 | -1.115 | 0.26491 |
| contrast14 | -1.52932 | 0.47978 | -3.188 | 0.00143** |
| contrast23 | -1.36377 | 0.48665 | -2.802 | 0.00507** |
| contrast24 | -0.76304 | 0.49583 | -1.539 | 0.12383 |
| contrast34 | 0.67017 | 0.62198 | 1.077 | 0.28127 |
| LanguageMan:contrast13 | 0.77349 | 0.81703 | 0.947 | 0.34379 |
| LanguageMan:contrast14 | 1.01836 | 0.73799 | 1.38 | 0.16761 |
| LanguageMan:contrast23 | 0.18709 | 0.71351 | 0.262 | 0.79316 |
| LanguageMan:contrast24 | 0.06119 | 0.72168 | 0.085 | 0.93243 |
| LanguageMan:contrast34 | -0.45711 | 0.89534 | -0.511 | 0.60967 |

* = denotes significance, contrast13 indicates contrast T1-T3, contrast14 indicates T1-T4, etc.

Additionally, in order to compare all Mandarin contrasts to one another, a Tukey post-hoc test was carried out. The post-hoc test showed that listeners performed significantly better on the contrasts T1-T2 and T3-T4 than on T1-T4 and T2-T3. Table 2 illustrates the results of these planned comparisons.

**Table 2. Results of Tukey post-hoc planned comparisons for Mandarin tonal contrasts**

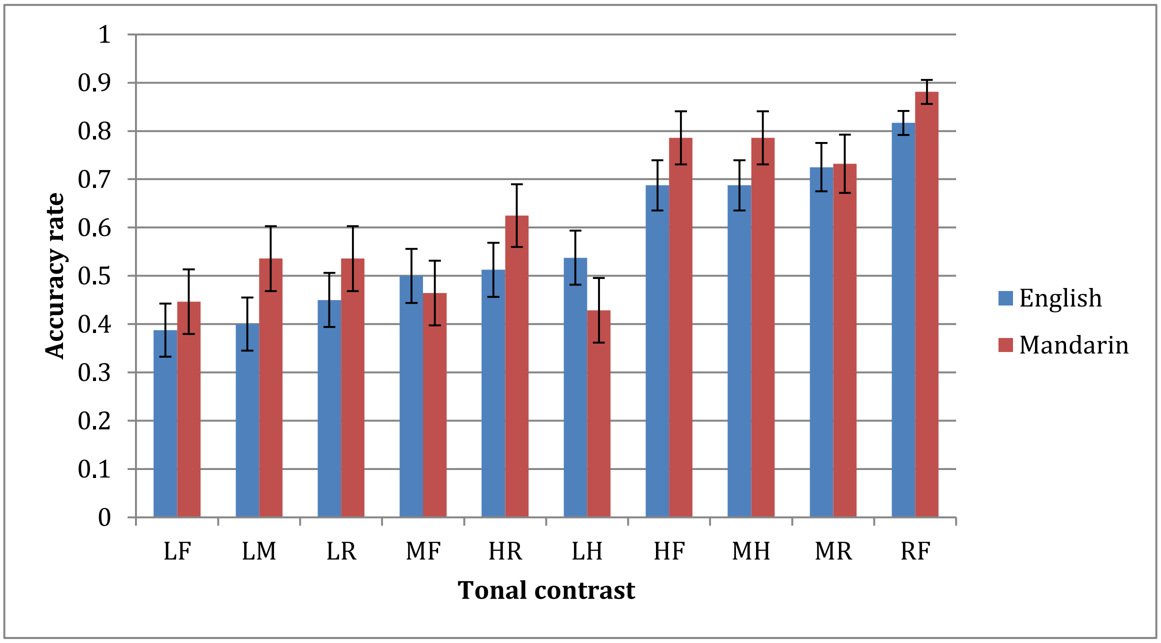
| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| 13 - 12 | -0.5617 | 0.5038 | -1.115 | 0.87231 |
| 14 - 12 | -1.5293 | 0.4798 | -3.188 | 0.01735* |
| 23 - 12 | -1.3638 | 0.4867 | -2.802 | 0.05543(*) |
| 24 - 12 | -0.763 | 0.4958 | -1.539 | 0.63276 |
| 34 - 12 | 0.6702 | 0.622 | 1.077 | 0.88767 |
| 14 - 13 | -0.9677 | 0.4261 | -2.271 | 0.20124 |
| 23 - 13 | -0.8021 | 0.4155 | -1.931 | 0.37692 |
| 24 - 13 | -0.2014 | 0.4436 | -0.454 | 0.99751 |
| 34 - 13 | 1.2318 | 0.5651 | 2.18 | 0.24179 |
| 23 - 14 | 0.1656 | 0.4221 | 0.392 | 0.99876 |
| 24 - 14 | 0.7663 | 0.3957 | 1.936 | 0.37333 |
| 34 - 14 | 2.1995 | 0.5612 | 3.919 | 0.00124** |
| 24 - 23 | 0.6007 | 0.4394 | 1.367 | 0.74166 |
| 34 - 23 | 2.0339 | 0.5547 | 3.667 | 0.00329** |
| 34 - 24 | 1.4332 | 0.5744 | 2.495 | 0.12208 |

* = denotes significance, (*) = marginally significant result

These results are consistent with previous researchers' findings, which taken together, suggest the most difficult contrasts to discriminate for English speakers are T1-T4 and T2-T3 (Hao, 2012; Mantle & Hoskin, 1997; So, 2006). The discrimination of T2-T3 is found to be the most difficult, and Hao (2012) proposes that there is an intrinsic difficulty associated with perceptual discrimination of those two tonal contrasts, having found Cantonese speakers to exhibit parallel difficulties in discrimination. One possible explanation for this phenomenon is that T3 in Mandarin is affected by tone sandhi, wherein T3 undergoes a phonological change to T2 before another T3 (T3-T3 sequences are phonetically realized as T2-T3). The phonological distinction is thus neutralized in such contexts, resulting in what Hume and Johnson (2003) call a “partial phonological contrast,” which perhaps inhibits the formation of distinct tone categories for T2 and T3. Tseng and Lee (2012) provide support for this theory in a study where Mandarin-learning English speakers were asked to participate in a tone identification task. They had the most difficulty identifying T2, which was mistaken for T3 in 73.4% of errors, probably because of the neutralization of the phonemic contrast in T3-T3 sequences. Other acoustical cues, such as the temporal location of the turning point within the syllable, may be more salient in identifying T2 vs. T3 for native speakers(Shen & Lin, 1991).

3.2 Thai accuracy

**Figure 2. Accuracy response rates for Thai by tonal contrast.**




In analysis of Thai accuracy data, the reference levels are English monolingual for the Language factor and LM contrast for the Contrast factor. Like the results of the accuracy data for Mandarin tonal contrasts, there was again no significant effect of language (ß = 0.521, z = 1.234, p = .217). The Mandarin learners performed no better or worse than the Mandarin-naïve participants. As shown in table 3, there are some significant effects of contrast when LM is taken as the fixed reference point: HF, MH, MR, and RF all differ significantly from LM. LM was chosen as the reference for this comparison following Abramson's (1976) observation that this contrast was the most difficult for his participants, a finding that is replicated here. There are no significant interactions, except for LH, which is one of two contrasts (the other being MF) where the English group achieves better accuracy than the Mandarin learners.

One possible explanation for the inconsistencies in response accuracy rates could be presentation of irregular stimuli, so stimuli were re-examined to investigate this unexpected result. Evidence of creaky voice would be particularly relevant, especially for the level stimuli, for which glottalization has been suggested to serve as a possible source of phonemic information (Abramson, 1976; Vance, 1977), however no perceptual abnormalities in the recorded speech were observed.

**Table 3. Results of mixed-effects binomial logistic regression for Thai tonal contrasts**

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | -0.39381 | 0.3397 | -1.159 | 0.24635 |
| LanguageMan | 0.587 | 0.37395 | 1.57 | 0.11648 |
| contrastHF | 0.97332 | 0.35575 | 2.736 | 0.00622** |
| contrastHR | 0.2821 | 0.36126 | 0.781 | 0.43488 |
| contrastLF | 0.12344 | 0.40666 | 0.304 | 0.76148 |
| contrastLH | 0.19318 | 0.40376 | 0.478 | 0.63233 |
| contrastLR | 0.69883 | 0.38663 | 1.807 | 0.07069 |
| contrastMF | 0.6293 | 0.36881 | 1.706 | 0.08795 |
| contrastMH | 1.00153 | 0.39635 | 2.527 | 0.01151* |
| contrastMR | 1.11289 | 0.38523 | 2.889 | 0.00387** |
| contrastRF | 2.19312 | 0.33038 | 6.638 | < 0.001*** |
| LanguageMan:contrastHF | -0.06362 | 0.54474 | -0.117 | 0.90703 |
| LanguageMan:contrastHR | -0.10911 | 0.50776 | -0.215 | 0.82986 |
| LanguageMan:contrastLF | -0.31533 | 0.51588 | -0.611 | 0.54103 |
| LanguageMan:contrastLH | -1.0987 | 0.52032 | -2.112 | 0.03472* |
| LanguageMan:contrastLR | -0.16899 | 0.52513 | -0.322 | 0.7476 |
| LanguageMan:contrastMF | -0.7788 | 0.539 | -1.445 | 0.14848 |
| LanguageMan:contrastMH | -0.06398 | 0.54421 | -0.118 | 0.90641 |

| | | | | |
|---|---|---|---|---|
| LanguageMan:contrastMR | -0.55427 | 0.53409 | -1.038 | 0.29937 |
| LanguageMan:contrastRF | -0.02801 | 0.47266 | -0.059 | 0.95274 |

* = denotes significance

In order to compare all contrasts to one another, a Tukey post-hoc test was carried out. The post-hoc test showed that RF was significantly different from all other contrasts (see Table 4). No other comparisons reached significance, and they are not included in Table 4.

**Table 4. Results of Tukey post-hoc planned comparisons for Thai tonal contrasts**

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| RF - HF | 1.21891 | 0.33581 | 3.63 | 0.00994** |
| RF - HR | 1.91065 | 0.31455 | 6.074 | < 0.001*** |
| RF - LF | 2.06867 | 0.32239 | 6.417 | < 0.001*** |
| RF - LH | 1.99991 | 0.32875 | 6.083 | < 0.001*** |
| RF - LM | 2.19327 | 0.33038 | 6.639 | < 0.001*** |
| RF - LR | 1.49354 | 0.32079 | 4.656 | < 0.001*** |
| RF - MF | 1.56392 | 0.34089 | 4.588 | < 0.001*** |
| RF - MH | 1.19103 | 0.32894 | 3.621 | 0.01057* |
| RF - MR | 1.0807 | 0.34114 | 3.168 | 0.0485* |

* = denotes significance, (*) = marginally significant result

To compare these results to Schaefer and Darcy (2013), the tone contrasts were also collapsed into three categories: contour (RF), level (LM, MH, LH), and mixed (LR, LF, MR, MF, HR, HF). Lee and Nusbaum (1993) suggest that perception of suprasegmentals is different for contour vs. level tones, and that Mandarin speakers attend to level tones integrally with speech in contrast to English speakers, who only do so for the Mandarin contoured tonal contrasts included in the study.

**Figure 3. Accuracy response rates for Thai tonal contrasts by category**

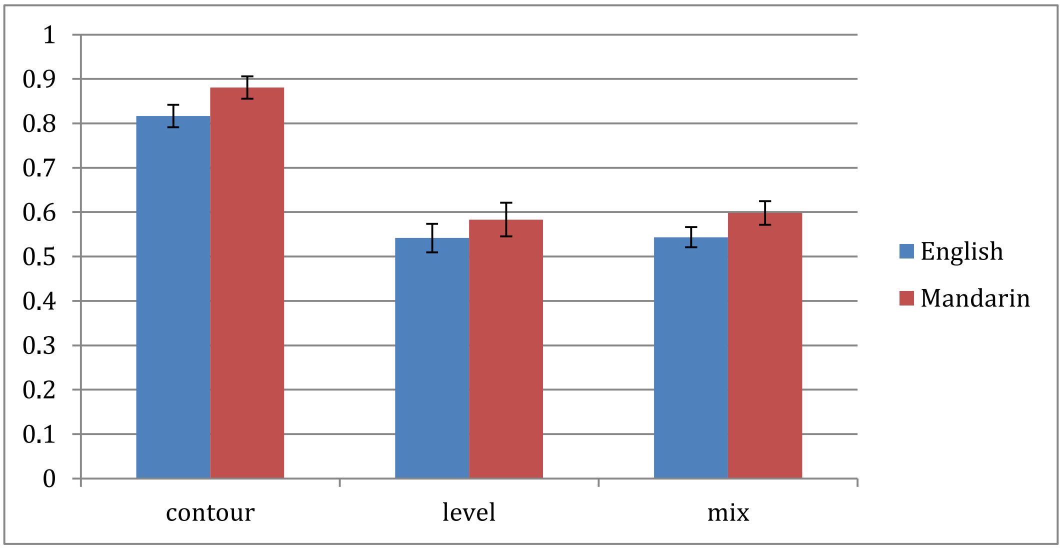



**Table 5. Results of mixed-effects linear regression for Thai tonal contrasts by category**

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 1.8644 | 0.2919 | 6.388 | < 0.001*** |
| LanguageMan | 0.5641 | 0.3232 | 1.745 | 0.081(*) |
| tonetypelevel | -1.8636 | 0.2319 | -8.038 | < 0.001*** |
| tonetypemix | -1.6873 | 0.2118 | -7.968 | < 0.001*** |
| LanguageMan:tonetypelevel | -0.3758 | 0.374 | -1.005 | 0.315 |
| LanguageMan:tonetypemix | -0.3032 | 0.3445 | -0.88 | 0.379 |

* = denotes significance, (*) = marginally significant result

From this perspective, there is a marginally significant effect of language for contour tonal contrasts (RF), and the accuracy rate for those contrasts differ significantly from the accuracy rates for level and mixed contrasts. Tukey post-hoc comparisons confirm that accuracy rates for the level and mixed contrasts are significantly lower than the accuracy rate for contour contrasts.

**Table 6. Results of Tukey post-hoc planned comparisons for Thai tonal contrasts by category**

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| level - contour | -1.8636 | 0.2319 | -8.038 | < 0.0001*** |
| mix - contour | -1.6873 | 0.2118 | -7.968 | < 0.0001*** |
| mix - level | 0.1763 | 0.1745 | 1.01 | 0.567 |

* = denotes significance

This result contrasts with Schaefer and Darcy (2013), whose Mandarin, Japanese, and Korean listeners all displayed the above pattern of results, with contour contrasts as the most discriminable. Their English speakers, on the other hand, performed best on mixed contrasts, followed by direction, and then height. The pattern displayed here suggests that perhaps the English speakers tested by Schaefer and Darcy were an anomaly.

**4. Discussion**

Although the hypothesis that Mandarin learners would perform better on both Mandarin and Thai tonal contrasts than naïve American English speakers was not supported, patterned differences in perception of tonal contrasts provide insight into the processes behind tonal language acquisition for non-tone language speakers.

The observation that participants performed best on the Mandarin T1-T2 and T2-T4 contrasts and most poorly on T1-T4 and T2-T3 contrasts, and that they performed best on contrasts involving the rising and falling (RF) Thai tones, supports the hypothesis that differences between contoured and level tones affect their intrinsic perceptibility (Abramson, 1977; Chiao et al., 2011; Francis et al., 2003; Y. Lee et al., 1996; W. S. Wang, 1976). Using subtly manipulated stimuli, Gårding et al. (1986) find that in perception of T3 and T4, it is a *change from* falling or *change to* falling that is integral for accurate perception. In the present study, contrasts for Mandarin T1-T2 and T2-T4 involve tones rising or falling in opposite directions, and Thai contour contrasts consist of only the RF distinction. It is clear that a perceptual advantage is conferred when contrasts involve tones moving in opposite directions. Comparing this study to others involving native speakers of languages with multiple level lexical tonal contrasts provides further support to the hypothesis that native-language influence has a strong effect on perception of suprasegmental tone.

In a study by Chiao et al. (2011), effects of L1 interference on perception of non-native level tonal contrasts were investigated. Participants were presented with an ABX task testing their discrimination of Toura level tonal contrasts, a Niger-Congo language containing 4 contrastive level tones. German (non-tone language) listeners performed significantly better than both Vietnamese (1 level tone) and Taiwanese (2 level tones) listeners, and the difference between the Vietnamese and Taiwanese groups was marginally significant ($p = 0.057$), with Taiwanese listeners performing the worst of all. This study provides strong evidence for perceptual assimilation of level tones to native categories, in contrast to the non-categorical mode of perception that Abramson (1976) and Francis et al. (2003) predict for level tones. In Chaio et al.'s experiment, German listeners performed best as they failed to assimilate stimuli to native-language tone categories. Vietnamese listeners then performed second-best and significantly above chance, probably because they are natively trained to attune to suprasegmental differences in speech, like the Mandarin speakers of Lee and Nusbaum's (1993) study. However, since they possess only one level tone in their native phonology, all Toura tones are perceived with CG assimilation to the same category. Taiwanese speakers performed the worst; with two level tones in their native language, Toura tones that were close in phonological space to Taiwanese tones may have been erroneously assimilated to a single category (SC) with less variance in category goodness relative to the Vietnamese speakers, adversely affecting Taiwanese perception of Toura tonal contrasts. Other possible assimilation patterns for Taiwanese speakers include TC or CG, which would confer more accurate response rates than SC assimilation; however overall, the prediction that Taiwanese speakers would perform worse holds true. The marginal difference between the two tone language groups suggests that more research is necessary to disentangle the complicated relationship between L1 tonal categories and nonnative speech perception, however the consistent discrepancy between tone and non-tone language speakers suggests that L1 effects may be strong enough to confound effects of nonnatively learned tonal contrasts.

For the Mandarin learners tested in the present study, it appears that learning of lexical tone was not enough to form robust perceptual categories, even if they did display a numerical trend towards greater accuracy on tonal contrasts. Evidence from numerous studies does not reveal significant differences in perception of L2 sounds past a certain point as learners gain experience in a new language, leading Best and Tyler (2007) to propose that a cutoff to distinguish “experienced” from “inexperienced” learners be set no higher than 6-12 months of experience. By this definition, all of the learners in the present study would be classified as experienced learners, and performance on the perceptual task would not be expected to improve even with increased length of study or exposure to Mandarin. Since they do not show an advantage over naïve English speakers at this point, it is possible that Mandarin learners usually fail to improve their perceptual ability at distinguishing Mandarin tonal contrasts to the point of significance.

Although experienced Mandarin learners do not exhibit a significant perceptual advantage over naïve English speakers, auditory training studies have shown that American English speakers can improve significantly in perception of Mandarin tonal contrasts. Wang, Spence, and Sereno (2003) constructed a high-variability training program that conferred a 21% improvement in eight 40-minute sessions, demonstrating that English speakers can learn to perceive Mandarin tone better than naïve listeners. Additionally, participants were tested for retention 6 months later and found to exhibit insignificant attrition in perceptual abilities. Although some participants in the present study were highly advanced speakers who had input from many speakers throughout their acquisition of Mandarin, others had taken classes with only a few native instructors as teachers, and may not have developed robust ability to discriminate Mandarin tones. Insignificant differences in performance on the task in this experiment could be due to confounding effects of participants' inability to generalize their learning to new speakers in the recorded stimuli. Lively et al. (1993) find that in segmental training, participants may fail to generalize to a new speaker if limited input is given, a problem that high-variability auditory training resolves.

Another possibly confounding effect of the current study was a lack of control for musical training, which has been found to correlate with a perceptual advantage in lexical tone discrimination (Alexander, Wong, & Bradlow, 2005; Delogu, Lampis, & Belardinelli, 2006). Musically inexperienced participants in Alexander et al.'s (2005) study had not taken over 3 years of musical instruction in their lives, and none in the last eight years, and in Delogu et al.'s (2006), they had had no formal musical instruction at all. In contrast, the naïve English participants in the current study had an average of 5.2 years of instruction in music, and the Mandarin learners had an average of 4.69 years. Participants who were extremely attuned to acoustic differences in pitch through musical training may have demonstrated a confounding advantage in the perceptual task here.

No previous research involving an ABX discrimination task of Mandarin tonal contrasts for naïve American English speakers has been found with which to directly compare results, however results from a few previous studies may shed light on the relative accuracy rates of the participants in the current one. Alexander et al. (2005) found naïve American English participants to discriminate Mandarin tones with 71% overall accuracy in an AX task, in comparison to 81% in the present study's ABX task (mean of both naïve English speakers' and Mandarin learners' accuracy rates on all tonal contrasts). In contrast, their native Mandarin speakers discriminated tones with 89% accuracy and their American musicians discriminated with 87% accuracy. Falling right between the naïve listeners and musicians from Alexander et al. (2005), the participants in the current study represent a group with varying levels of musical experience.

Wang et al.'s (1999) study may also serve as a valuable comparison to the present one. Their pre-training Mandarin learners' worst accuracy rate was on the T2-T3 Mandarin contrast, with 75% accuracy in a tone identification task. All other contrasts were identified with 90% accuracy or better. Participants in the current study again performed comparably. Testing more naïve American English listeners on Mandarin tonal contrasts using natural speech stimuli and an ABX design would allow for a better baseline for comparison to the current study.

In another training study by Wayland and Guion, (2004) native Mandarin speakers were compared to American English speakers in discriminating and identifying the Thai LM contrast.

Mandarin speakers demonstrated significantly better performance than English speakers in the pre-test, however only the Mandarin group improved after a one-week training period. Their performance was, however, still significantly worse than the near-perfect native Thai speakers', indicating that their improvement did not correspond with development of new phonetic categories on par with native speakers'. This does suggest that English speakers have significantly more difficulty learning nonnative phonemic tonal contrasts than native speakers of a different tone language.

Additionally, the present discrimination task involved ABX stimuli that mixed two different speakers with different F0 vocal ranges. Moore and Jongman (1997) found that when they spliced tokens from a high-range speaker into a precursor phrase spoken by a low-range speaker or vice versa, tones were misidentified as listeners normalized for the previous speakers' vocal range. Abramson (1977) gave native Thai speakers an identification task using a continuum of simulated level tones. The ranges he obtained for level tone categories are reported in Table 7, beside the mean F0 values for the tokens used in this study.

**Table 7. F0 ranges of Abramson (1977) as well as mean F0 of the stimuli used in the present study.**

| | Abramson 1977 | Male Thai speaker mean F0 | Female Thai speaker mean F0 |
|---|---|---|---|
| Low | **< 108 Hz** | **134.95 Hz** | **177.85 Hz** |
| Mid | **108-133 Hz** | **142.85 Hz** | **193.20 Hz** |
| High | **> 133 Hz** | **174.61 Hz** | **225.08 Hz** |

Abramson's posited F0 ranges for each of the Thai level tones lie markedly below the F0 ranges of both of the speakers who recorded the stimuli used in the present study. The complete lack of overlap between the F0 ranges of the male and female speakers suggests that any level tone spoken by the male speaker might have been perceived as low relative to the previously presented female speech. The unusual reversal in accuracy rates for the LH distinction may be attributable to the similarities in F0 for the male high tone and female low tone.

Another explanation for the poor discrimination exhibited by the participants in this study may come from an extension of the original model of Best's PAM. Australian English and French speakers have demonstrated categorization of lexical tone to prosodic intonation categories (i-Categories) such as *statement*, *question*, and *exclamation* (So & Best, 2011). A clear pattern emerges wherein the non-tone language speakers place every type of tone into the "statement" i-Category, as illustrated in Table 8. If we extend PAM to suprasegmentals, we would predict that because every tone pattern is placed into the same category (SC assimilation), discrimination would be poor to moderate depending on the goodness of fit to the category.

**Table 8. Significant results of i-Categorization for Australian English and French speakers (So & Best, 2011)**

| | Australian English speakers' i-Categorizations | French speakers' i-Categorizations |
|---|---|---|
| Tone 1 | **statement, question** | **statement, exclamation** |
| Tone 2 | **statement, question** | **statement, question** |
| Tone 3 | **statement, question** | **statement** |
| Tone 4 | **statement, exclamation** | **statement, exclamation** |

For Australian English speakers, the only tone categorized differently was T4, which has a falling contour rather than rising or level. For French speakers, however, T1 and T4 are categorized identically, which is in line with findings that it constitutes one of the most difficult perceptual contrasts. If we look at the overall shape of each tone, the most comparable to Thai are T4 to falling and T3 to rising. This assimilation pattern has been observed in native Mandarin speakers who were naïve to Thai at the time of testing (Wu, Munro, & Y. Wang, 2011). In the current study, T3-T4 and RF each have the highest accuracy rating relative to other contrasts in Mandarin and Thai, respectively. Generalizing the findings of So and Best (2001) to Thai, the high discriminability of the Thai RF contrast may be attributable to TC perceptual assimilation to distinct i-Categories. Studies using American English speaking participants to investigate i-Category assimilation would be necessary to robustly compare results of the present study to So and Best, however general patterns in nontone language speakers' perceptions can be inferred from their findings.

The number of times that a tone was categorized as belonging to an "unknown" or unassimilable native i-Category was never significant, indicating that all the speech stimuli presented to Australian and French speakers were assimilable to familiar categories. The perception of tone has been compared to the vowel space in that acoustic differences occur on a continuum rather than discretely articulated as is the case with contrasts such as stop consonants at different points of articulation (Abramson, 1976). The assimilation of tones to native categories and apparent goodness of fit demonstrated by the lack of significance for "unknown" responses, suggests that like vowels for L1-learning infants, tones that are perceived as "good" exemplars are more difficult to discriminate than "poor" exemplars of native category prototypes (Grieser & Kuhl, 1989).

**5. Conclusions**

Although the Mandarin learners in this study had a numerical trend towards higher accuracy ratings on Mandarin and Thai tonal contrasts, the effect of adult L2 acquisition was insignificant overall. The lack of a perceptual advantage for Mandarin learners may be because the effects of learning only a nontone language in infancy disallow for the categorical discrimination of tonal contrasts that native tone language speakers exhibit for contoured tones. While native Mandarin speakers exhibit better perception of Thai tonal contrasts than native English speakers, Mandarin learners' perception was no better than that of Mandarin-naïve English speakers. Factors such as lack of control for musical experience, the inability to generalize tonal acquisition to new speakers, or the inability to account for the different F0 ranges of the male and female speakers in the stimuli recordings could have confounded the accuracy rates obtained. Examining the relative accuracy rates for different tonal contrasts, it seems evident that those involving level tones are significantly more difficult than contrasts that vary markedly in overall contour shape. Replicating the experiment with a larger set of participants, and controlling for the suggested confounding factors, may yield more significant results. Future research investigating the mechanisms involved in second language phonological category formation, as well as the relationship between intonation and tone, may provide useful in attempting to further understand the mechanisms at work in perceiving nonnative tonal contrasts.

References

Abramson, A. S. (1976). Thai tones as a reference system. *Tai Linguistics in Honor of Fang-Kuei Li*, 1-12.

Abramson, A. S. (1977). Noncategorical perception of tone categories in Thai. *The Journal of the Acoustical Society of America, 61*(S1), S66-S66. doi:http://dx.doi.org/10.1121/1.2015837

Alexander, J. A., Wong, P. C., & Bradlow, A. R. (2005, September). Lexical tone perception in musicians and non-musicians. In *INTERSPEECH* (pp. 397-400).

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes.

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook, 3*(1), 5-70.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words, 167*, 224.

Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.*

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America, 109*(2), 775-794.

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.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, 13-34.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., & Keller, P. (1996, October). Perception of lexical tone across languages: Evidence for a linguistic mode of processing. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on* (Vol. 4, pp. 2514-2517). IEEE.

Burnham, D., Kirkwood, K., Luksaneeyanawin, S., & Pansottee, S. (1992). Perception of central Thai tones and segments by Thai and Australian adults. *Pan-Asiatic Linguistics. Proceedings of the 3rd International Symposium of Language and Linguistics, Chulalongkorn University Press, Bangkok*, 546-560.

Chiao, W., Kabak, B., & Braun, B. (2011). *When more is less: Non-native perception of level tone contrasts*. Bibliothek der Universität Konstanz.

Delogu, F., Lampis, G., & Belardinelli, M. O. (2006). Music-to-language transfer effect: may melodic ability improve learning of tonal languages by native nontonal speakers?. *Cognitive Processing*, 7(3), 203-207.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 233-277.

Flege, J. E., Schirru, C., & MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication, 40*(4), 467-491.

Flege, J. E., & Davidian, R. D. (1984). Transfer and developmental processes in adult foreign language speech production. *Applied Psycholinguistics, 5*(04), 323-347.

Flege, J. E., McCutcheon, M. J., & Smith, S. C. (1987). The development of skill in producing word‐final English stops. *The Journal of the Acoustical Society of America, 82*(2), 433-447.

Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non) categorical perception of lexical tones. *Perception & Psychophysics, 65*(7), 1029-1044.

Gandour, J. (1984). Tone dissimiliarity judgments by Chinese listeners. *Journal of Chinese Linguistics, 12*(2), 235-261.

Gandour, J., Petty, S. H., & Dardarananda, R. (1988). Perception and production of tone in aphasia. *Brain and Language, 35*(2), 201-240.

Gårding, E., Kratochvil, P., Svantesson, J., & Zhang, J. (1986). Tone 4 and tone 3 discrimination in modern standard Chinese. *Language and Speech, 29*(3), 281-293.

Grieser, D., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology, 25*(4), 577.

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.

Hallé, P. A., Chang, Y., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32*(3), 395-421.

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

Hume, E., & Johnson, K. (2003, August). The impact of partial phonological contrast on speech perception. In *Proceedings of the fifteenth international congress of phonetic sciences* (pp. 2385-2388).

Iverson, P., & Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism?. *Perception & Psychophysics, 62*(4), 874-886.

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

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

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259-274) Springer.

Kuhl, P. K., & Iverson, P. (1995). Chapter 4: Linguistic experience and the "Perceptual magnet effect." *Speech Perception and Linguistic Experience: Issues in Cross-Language Research,* 121-154.

Lee, C. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech, 50*(1), 101-123.

Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics, 53*(2), 157-165.

Lee, Y., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research, 25*(5), 527-542.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94*(3), 1242-1255.

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

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

Mantle, D. R., & Hoskin, J. K. (1997). Tone acquisition in beginning mandarin learners. *Deseret Language and Linguistic Society Symposium, 23*(1) 65-80.

Massaro, D. W., & Cohen, M. M. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics, 13*(2).

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

Miller, J. L. (1978). Interactions in processing segmental and suprasegmental features of speech. *Perception & Psychophysics, 24*(2), 175-180.

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America, 102*(3), 1864-1877.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 8*(2), 297.

Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *The Journal of the Acoustical Society of America*, *97*(2), 1286-1296.

Polka, L., & Bohn, O. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America, 100*(1), 577-592.

Qin, Z., & Mok, P. (2012). The perception of speech and non-speech tones by tone and non-tone language listeners. *Proceedings of Speech Prosody*, 366-369.

Repp, B. H., & Lin, H. (1990). Integration of segmental and tonal information in speech perception: A cross-linguistic study. *The Journal of the Acoustical Society of America, 87*(S1), S46-S46. doi:http://dx.doi.org/10.1121/1.2028239

Schaefer, V., & Darcy, I. (submitted 2013). *Functional prominence of lexically-contrastive pitch in L1 shapes cross-linguistic perception of thai tones.* (Unpublished Indiana University)

Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech, 34*(2), 145-156.

So, C. K. (2006). Perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Proceedings of the 11th Australian International Conference on Speech Science & Technology*, 438-443.

So, C. K., & Best, C. T. (2011). Categorizing Mandarin tones into listeners' native prosodic categories: The role of phonetic properties. *Poznań Studies in Contemporary Linguistics PSiCL*, *47*, 133.

Tseng, C., & Lee, A. (2012). An online perceptual test of Mandarin Chinese tones by L2 learners of tonal vs. nontonal languages. *Tonal Aspects of Languages-Third International Symposium.*

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

Vance, T. J. (1976). An experimental investigation of tone and intonation in Cantonese. *Phonetica*, *33*(5), 368-392.

Vance, T. J. (1977). Tonal distinctions in Cantonese. *Phonetica*, *34*(2), 93-107.

Wang, W. S. (1976). Language change*. *Annals of the New York Academy of Sciences*, *280*(1), 61-72.

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

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning, 54*(4), 681.

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America, 75*(6), 1866-1878.

Wood, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *The Journal of the Acoustical Society of America, 60*(6), 1381-1389.

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. (2011). The Effects of linguistic experience on perceptual assimilation of lexical tone. *Proceedings of ICPhS XVII*, 17-21.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*(1), 61-83.

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.

**Appendix**

### Table 9. List of Thai experimental triads

| order | contrast | item |
|---|---|---|
| ABA | LM | naa |
| ABB | LM | pii |
| BAA | LM | chaa |
| BAB | LM | dii |
| ABA | LH | huu |
| ABB | LH | kai |
| BAA | LH | maa |
| BAB | LH | mai |
| ABA | MH | mii |
| ABB | MH | baa |
| BAA | MH | bpuu |
| BAB | MH | ruu |
| ABA | LR | suu |
| ABB | LR | ruu |
| BAA | LR | waa |
| BAB | LR | yaa |
| ABA | LF | kai |
| ABB | LF | huu |
| BAA | LF | mii |
| BAB | LF | suu |
| ABA | MR | bpuu |
| ABB | MR | baa |
| BAA | MR | naa |
| BAB | MR | ruu |
| ABA | MF | chaa |
| ABB | MF | mai |
| BAA | MF | pii |
| BAB | MF | waa |
| ABA | HR | dii |
| ABB | HR | maa |
| BAA | HR | ruu |
| BAB | HR | yaa |
| ABA | HF | naa |
| ABB | HF | pii |
| BAA | HF | ruu |
| BAB | HF | yaa |
| ABA | RF | baa |
| ABB | RF | huu |
| BAA | RF | mii |
| BAB | RF | suu |

| | | |
|---|---|---|
| ABA | RF | bpuu |
| ABB | RF | kai |
| BAA | RF | ruu |
| BAB | RF | chaa |
| ABA | RF | maa |
| ABB | RF | waa |
| BAA | RF | dii |
| BAB | RF | mai |

**Table 10. List of Mandarin experimental triads**

| order | contrast | item |
|---|---|---|
| ABA | 12 | pou |
| ABB | 12 | fei |
| BAA | 12 | cha |
| BAB | 12 | ni |
| ABA | 13 | fa |
| ABB | 13 | cha |
| BAA | 13 | ma |
| BAB | 13 | ni |
| ABA | 14 | tou |
| ABB | 14 | cha |
| BAA | 14 | fei |
| BAB | 14 | ma |
| ABA | 23 | mou |
| ABB | 23 | ni |
| BAA | 23 | fa |
| BAB | 23 | cha |
| ABA | 24 | na |
| ABB | 24 | ma |
| BAA | 24 | fei |
| BAB | 24 | cha |
| ABA | 34 | fa |
| ABB | 34 | cha |
| BAA | 34 | ni |
| BAB | 34 | ma |

**Demographic information Questionnaire**

1. **Code**

   Day of the month of this experiment (2 numbers): .

   Day of birth (2 numbers):

   First two letters of your hometown:

2. **Age:**

3. **Where have you lived, and for how long? Start with your hometown.**

4. **Gender:** male female

5. **First/native language(s):**

6. **Second/other language(s) spoken, age of acquisition, and ability (fluent, good, mostly reading, a little, etc):**

7. **Language comments** (e.g. you do not use first language any more, started school/reading in non-native language)**:**

**8. Have you ever been diagnosed with a speech disorder or impediment? If so, please explain.**

**9. Do you have any formal or informal musical training? Please indicate in what (voice, instrument, theory), how old you were when you began study, and for how long.**