DISCRIMINATION OF TONE CONTRASTS IN MANDARIN DISYLLABLES BY NAÏVE AMERICAN ENGLISH LISTENERS

by

SHARI SALZHAUER BERKOWITZ

A dissertation submitted to the Graduate Faculty in Speech-Language-Hearing Sciences
in partial fulfillment of the requirements for the degree of Doctor of Philosophy,
The City University of New York

2010

This manuscript has been read and accepted for the

Graduate Faculty of the Speech-Language-Hearing Program in satisfaction of the

dissertation requirement for the degree of Doctor of Philosophy

_____	_____
Date	Winifred Strange, Ph.D. Chair of Examining Committee
_____	_____
Date	Klara Marton, Ph.D. Executive Officer

Supervisory Committee:

Winifred Strange, Ph.D.

Valerie Shafer, Ph.D.

Brett Martin, Ph.D.

THE CITY UNIVERSITY OF NEW YORK

Abstract

DISCRIMINATION OF TONE CONTRASTS IN MANDARIN DISYLLABLES BY NAÏVE AMERICAN ENGLISH LISTENERS

by

Shari Salzhauer Berkowitz

Advisor: Winifred Strange, Ph.D.

The present study examined the perception of Mandarin disyllabic tones by inexperienced American English speakers. Participants heard two naturally-produced Mandarin disyllables, and indicated if the two were the same or different. A small native Mandarin-speaking control group participated as well. All 21 possible Mandarin contrasts where the initial syllable varied but the final syllable stayed the same were tested. Acoustic analysis was performed on the stimuli under study. Mandarin subjects scored at ceiling on all contrasts. American English subjects performed poorly on contrasts where the difference in mean F0 was small, or where the difference in the offset F0 of the first syllable was small. They also performed poorly when the difference in slope of the final syllable was small. Previous research has proposed that American English listeners attend primarily to the height difference between two tone stimuli, but here they attended to height in the first syllable and contour in the second syllable.

Acknowledgements

My world has been filled with loving people who encouraged me to be inquisitive and to search for solutions. I want to thank a lifetime of family and friends, starting with my parents, and my aunt and uncles, continuing through many wonderful teachers, and culminating in my children. Your support has meant everything to me.

Winifred Strange has been a staunch supporter as well as a strict taskmaster on an as-needed basis. I will take her enthusiasm for some good, fresh data with me always. Her love of Science with a big "S" has been a framework under which all ideas can flourish.

The level of collaboration and care that has developed amongst the past and present members of the Speech Acoustics and Perception Lab has made it all possible for me on so many levels. Apparently not every lab is like that. I know that the spirit we shared will continue into new labs and new collaborations and I look forward to it all.

Thanks to committee members Valerie Shafer and Brett Martin, for your support and encouragement. It has been my pleasure to work with you.

Thanks also to Doug Whalen, who served as the Outside Reader. He provided not only insightful comments, but great cheer as well.

The CUNY Graduate Center Doctoral Student Research Grant Program, Competition #4 provided subject money, for which I am extremely grateful.

It has been quite a ride and I thank my friends for taking it with me. You know who you are.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1. Introduction

The field of cross-language speech perception examines how naïve listeners perceive speech sounds in an unfamiliar language, and how the listener's native language(s) impact on this perception. A large body of research has developed over the years, and continues to grow. Two oft-cited theories of cross language speech perception, the Perceptual Assimilation Model (Best, 1995) and the Speech Learning Model (Flege, 1995) attribute the difficulties experienced by listeners in an unknown language to the relationships between the inventories of consonant and vowel sounds in the known and the unknown languages. These models do not specifically address cross-language tone perception. When comparing speech sounds in the native language to speech sounds in the language under study, it is typically specified whether the sound occurs in the native language as an allophone, or not at all. However, this type of comparison for tone and non-tone languages is not really possible. All languages make use of fundamental frequency variation, although not all languages do so on a syllable-by-syllable basis as tone languages do. The present study generated new data regarding cross-language tone perception by presenting natural stimuli from a tone language to speakers of a non-tone language, and examined the relative difficulty of tone perception for naïve listeners. Naïve American English subjects listened to Mandarin disyllabic utterances whose tones varied on the first syllable, with all possible tone contexts in the second syllable. Participants judged whether two disyllabic strings had the same tone pattern or not. The current study is one of only a few studies that examined cross-language perception of Mandarin tones by American English (AE) speakers using disyllabic stimuli.

1.1 Characteristics of Mandarin Tones

The majority of the world's languages are tone languages, and Mandarin Chinese is the most widely spoken of these (Yip, 2002). In a tone language, the meaning of a string of speech sounds produced with one fundamental frequency (F0) pattern is different from the same string produced with a different F0 pattern. For example, in Mandarin, the word /yi/ means 'clothing' when produced with a high, level tone. In the most common notations system, pinyin, this high level pattern is labeled "Tone 1." It denotes 'aunt' when produced with a rising tone (Tone 2), 'chair' when produced with a low/dipping tone (Tone 3) and 'meaning' when produced with a falling tone (Tone 4). The four Mandarin tones differ not only in F0 height, but also in contour and timing (see Figure 1). When produced in isolation, Tone 1 is produced with a high F0 which stays relatively steady across the whole syllable. Tone 2 starts at a medium F0, drops slightly and turns toward rising after 20% of the vowel duration, and rises to about the same height as Tone 1 (Xu, 1997). Tone 3 starts out lower than Tone 2, drops a considerable amount, and then rises after 40% of the vowel duration when produced in isolation (Xu 1997). In connected speech in non-final position, Tone 3 is much less likely to rise, and can be considered a low, falling tone without any rise. The relative duration up to the turning point of the F0 contour may be the critical cue for native speakers to distinguish Tones 2 and 3 in isolation (Shen, Lin & Yan, 1993; Shen & Lin, 1991). Tone 4 starts even higher than Tone 1 and drops precipitously.

The four tones also show consistent differences in duration when produced in isolation, with Tone 3 being the longest and Tone 4 being the shortest (Xu, 1997).

Finally, Mandarin tones vary in intensity, with Tone 3 being the least intense and Tone 4 being the most intense (Whalen & Xu, 1992).

The primary perceptual cue to a tone's identity is its F0 contour (Luo & Fu, 2004, Yip 2002), but native speakers are able to use the covarying amplitude and duration cues in degraded listening conditions. For example, when the F0 information was replaced with noise, native speakers were able to use naturally occurring amplitude envelopes and durational information to identify the tones presented in isolation (Whalen & Xu, 1992). Duration alone is the least informative cue to tone identity (Fu & Zeng, 2000). It is not known whether AE speakers attend primarily to F0, or whether they rely on amplitude or duration when discriminating Mandarin tone pairs.

1.2 Tone coarticulation

Previous cross-language studies of lexical tone perception have focused largely on monosyllables produced in isolation. However Mandarin words are primarily disyllabic (Duanmu, 1999) and, of course, most utterances are multisyllabic. When combining two syllables in Mandarin, there are 15 possible tone contours created[i] and, upon examination, it is difficult to describe the differences between disyllabic tone combinations as primarily height, primarily contour, or primarily timing of the turning point. Except for disyllables with Tone 1 followed by Tone 1 (hereafter Tone 11), all the disyllabic tone combinations are contour tones.

As the syllables of words and phrases are strung together into utterances, the tone of each syllable is coarticulated with its neighbors (Xu, 1994, 1997, 2001). In a disyllable, each syllable's F0 influences the other; however carry-over coarticulation changes the final F0 contour more than anticipatory coarticulation changes the first tone.

That is, the resultant surface form of the F0 pattern shows more influence by the earlier syllable on the later syllable than vice versa. Thus, when comparing two disyllables with different initial syllables but identical final syllables, such as Tone 21 vs. Tone 31, the surface forms of the Tone 1 syllables may be substantially different. Whereas a native speaker would be expected to easily identify these two terminal syllables as having the same tone, a naïve listener without the appropriate language experience might misinterpret this systematic variation as an additional meaningful difference. Thus, coarticulation may introduce additional variability to which the non-native listeners might attend. (The actual F0 tone contours used in this study are presented graphically in Appendix G, Figure G2. Tone 31 is a good example of the surface form of the final syllable deviating from the monosyllabic form, because it is neither high nor level.)

The coarticulatory effects of combining two syllables can also be analyzed in terms of the compatibility of the two syllables (Xu, 1997). If the F0 offset of the first tone is markedly different from the F0 onset of the second tone, this is known as a non-compatible context, and this context will alter the contours from their canonical form more than if the two tones are compatible. For example, Tone 14 is considered compatible; the offset of Tone 1 is high, and the onset of Tone 4 is high. But Tone 12 is non-compatible; the first tone ends at a high F0 but the second tone is rising to high. In order to rise, it must first fall. Thus, the final form of the non-compatible Tone 12 is less like its two canonical tones abutted than the final form of Tone 14. Again, an experienced listener would know that the resultant surface differences are not relevant to the final syllables' tone category, but a naïve listener might not. Tones in compatible contexts have more extreme slopes, whereas tones in non-compatible contexts show

decreased slope excursion (Xu, 1994). Thus, pairs of disyllables whose second syllables are the same tone will have similar slopes when both preceding syllables are of the same compatibility. Pairs of disyllables whose second syllables are the same tone but whose first syllables are of opposite compatibility will show a larger difference in the slopes of the second syllable.

Another way to conceptualize this can be addressed by the example in Appendix G, figure G2, panel 2. Initial Tones 3 and 4 end at a low F0. Tone 2 starts at a low F0 and rises. Thus, these combinations are considered compatible, indicated on the figure as a "+." However, Tones 1 and 2 end at a high F0. In order to rise for Tone 2, they must first begin to fall. So, these combinations are considered non-compatible, indicated on the figure as a "-." It can be seen that the gap in frequency height, and the difference in frequency contour are more pronounced when one curve is compatible and one is non-compatible.

1.3 Neurobiology of tone processing

A variety of neurobiological methods have been used to explore whether tone and non-tone language speakers have different patterns of brain activity in response to linguistic tone (Wang, Jongman & Sereno, 2006). Language background does not affect the processing of pitch glides (Bent, Bradlow & Wright, 2006), but it does affect the processing of linguistic tones. Thus, language experience changes the brain's neural circuits for tone-related language processing.

It has been claimed that speakers of tone languages process lexical tones primarily in the left cerebral hemisphere, and non-tone speakers process the same stimuli primarily in the right cerebral hemisphere (Wang, Jongman & Sereno, 2006). However,

in most studies of lexical tone processing, the Mandarin subjects are hearing real words, whereas the AE subjects are hearing multiple productions of one word, but not lexically contrastive words. For example, two studies used the stimulus set yi1, yi2 and yi3 (Chandrasekaran, Gandour & Krishnan, 2007; Chandrasekaran, Krishnan & Gandour, 2007). For a Mandarin speaker, these are interpreted as 'clothing,' 'aunt' and 'chair.' To an AE listener, all three productions would likely sound close to /i/, glossing to 'E,' but with three different intonation contours. This is a possibly confounding issue, because the left hemisphere is typically associated with phonological processing, and the right with prosodic and melodic processing. This difference in laterality has been shown with various methods, including the mismatch negativity electrophysiological response (Chandrasekaran, Gandour & Krishnan, 2007; Chandrasekaran, Krishnan & Gandour, 2007) and Positron Emission Tomography (Gandour, et.al., 2000; Hsieh, Gandour, Wong & Hutchins, 2001; Klein, Zatorre, Milner & Zhao, 2001). Even processing at the brain stem level appears to be affected by language experience; Mandarin speakers showed a stronger Frequency Following Response (FFR) than AE speakers in response to synthetic time- and intensity-normalized Mandarin tones (Krishnan, Xu, Gandour & Cariani, 2005). The FFR is a measure of neural phase locking over time (Krishnan, 2007) and can be thought of as a measure of efficient neural entrainment. This effect was not seen when the same populations were tested with non-speech tone analogs designed to have linear slopes and identical onset and offset frequencies as the lexical tone stimuli (Xu, Krishnan & Gandour, 2006). The advantage was only seen when speakers of tone languages listened to speech-like stimuli. Similarly, in a two-alternative forced choice task, speakers of Mandarin did not show any better accuracy than speakers of American

English when they discriminated non-speech pitch glides (Bent, Bradlow & Wright, 2006).

Mandarin speakers have been shown to rely more heavily on the differences in tone contour, whereas non-tone speakers have been shown to rely primarily on differences in mean F0 height when perceiving lexical tones in isolation (Guion and Pederson, 2007; Kaan, Wayland, Bao & Barkley, 2007, Chandrasekaran, Gandour & Krishnan, 2007; Chandrasekaran, Krishnan & Gandour, 2007). Chandrasekaran, Krishnan & Gandour (2007) tested AE speakers and Mandarin speakers in a passive oddball ERP paradigm, and measured mismatch negativity (MMN), a pre-attentive measure of detection of change, in response to resynthesized duration- and intensity-normalized Mandarin tones. Tone 3 was presented as the oddball stimulus (15% of stimuli) with Tone 1 as the standard in one condition and Tone 2 as the standard in another condition. The amplitude of MMN responses by AE subjects to Tone 3 as an oddball was the same regardless of which standard was tested, but the Mandarin subjects showed a much larger amplitude MMN in the condition where Tone 1 was the standard than in the condition where Tone 2 was the standard; this response was also of greater amplitude than the AE subjects' MMN. That is, the Mandarin subjects show a much larger MMN response to the level tone vs. contour tone condition (Tone 1 vs. Tone 3) than to the two contour condition (Tone 2 vs. Tone 3). For the AE subjects, the conditions were not statistically significantly different—the height differences in both contrasts were sufficient to elicit a small MMN in both cases. This supports the argument that AE listeners are attuned to height differences and MA listeners to contour differences.

In a follow-up study, Chandrasekaran, Gandour & Krishnan (2007) also tested ten Mandarin and ten AE speaking subjects on the above two conditions, and added a condition in which Tone 1 was the standard and Tone 2 was the deviant. Data from the Fz electrode was analyzed for each of 20 participants, and each participant's MMN responses were entered into an individual matrix. In this way, the researchers determined which stimuli elicited large or small MMN for each participant. The resultant matrices were evaluated with INDSCAL multidimensional scaling. If the Mandarin subjects and the AE subjects were processing the tones in the same way, only one dimension would have been required to classify the data. However, two dimensions were required to classify the set into language groups. The authors interpret the dimensions used as height (measured by mean F0 and offset F0) and contour (measured by degree of slope, especially the slope after the turning point toward rising). Tone 4 was not tested and so it is unclear how including the falling contour would change the results of this study. Mandarin speakers responded on the basis of the contour dimension more than the height dimension; for the AE listeners, height was more important. This work has not yet been extended to disyllables.

Many different methods have been used in quantifying the contours of tones. Tsao (2008) measured the Turning Point (turning point = time at F0 minimum ÷ tone duration x 100%). Liu, Tsao and Kuhl (2007) used a measure of Relative Turning Point (relative turning point = (duration of onset to valley ÷ duration of vowel) x 100%). Other authors have proposed other metrics, but these measures were computed after speaker equalization (Wang, Jongman & Sereno, 2003) or were computed on three time-normalized stimuli (Chandrasekaran, Gandour & Krishnan, 2007). All the above

methods of quantifying the shape of the F0 contour have strengths and weaknesses, but all were computed on single syllable tones. Xu (1994) proposed that in multisyllabic stimuli, the "coefficient of a simple linear regression line" is a better measure than examining Turning Point. This method was used for acoustic analysis, discussed below in Chapter 2.

1.4 Perception of tone in multisyllabic stimuli

One feature of all the studies discussed so far is that they used monosyllabic stimuli. Some were produced naturally in citation form, some were recorded in sentences and then excised, some were synthesized or otherwise modified from naturally produced tokens, but all were monosyllabic. However, a few studies have begun to look at perception of multisyllabic stimuli. Gottfried and Suiter (1997) employed a subset of Mandarin disyllables in a study whose subjects were late learners of Mandarin with primarily classroom-based experience (mean 2.75 years). The subjects heard Mandarin disyllables di, da, duo, and du with all possible tones, followed by the carrier syllable /dzɨ/ produced with Tone 4. These stimuli were excised from naturally produced sentences. Subjects were asked to identify the stimuli on a 16-choice answer sheet (4 tones x 4 vowels). They found that initial syllables of disyllabic stimuli were identified less accurately than monosyllables by late learners of Mandarin, even when the second syllable remained the same across all the stimulus items. The study was then repeated with the final carrier syllable removed. The L2 Mandarin students identified the tone of the target syllable (in initial position) in context with 61% accuracy, whereas they identified the excised syllable with 77% accuracy. A native Mandarin speaking control group scored at ceiling in both conditions (99% and 97.5% respectively). So, even with

the same carrier syllable in every trial, subjects who were students of Mandarin found identification of the tone of the initial syllable of a disyllable more challenging than that of monosyllables.

Listeners learn to attend to linguistically relevant sources of information in the acoustic signal of their native language and ignore the irrelevant ones. Given that F0 contours are used at the phrase/utterance level to indicate linguistically-relevant intonation differences in English, when AE speakers listen to the F0 contour of disyllables, it is possible that they will process the pattern across the two syllables as one unit. Broselow, Hurtig & Ringen (1987) suggested that AE listeners were most accurate on identifying Tone 4 when it occurred in the final position of a disyllable or trisyllable. They proposed that the falling contour in the final position sounds most natural to AE listeners due to the tendency for English declarative sentences to have a falling F0 at the end. After a short training session with monosyllables in which the subjects learned to label monosyllables as Tone 1, 2, 3 or 4, they were given disyllabic and trisyllabic stimuli to label. AE speakers were 94% accurate on identifying Tone 4 in isolation, but only 31% accurate when Tone 4 was the initial syllable in a two syllable string. The authors propose that this is interference from English intonation, and supply additional anecdotal information that students of Mandarin describe Tone 4 as the only one that sounds "normal" to them.

If AE listeners are attending to Mandarin disyllables as a short phrase, it is possible that the acoustic information present in both syllables together is processed as a whole, rather than on a syllable-by-syllable basis. In this dissertation, the behavioral data from AE subjects was correlated with data from an acoustic analysis of disyllabic stimuli.

If subjects were treating the two syllables as one prosodic unit, this should be revealed by a stronger correlation between behavior and the mean difference in F0 on the whole disyllable, as opposed to the mean difference in F0 on just the initial, contrastive syllable.

1.5 Pilot study

Many of the studies discussed above found that AE listeners relied on F0 height as the primary cue to tone contrasts in monosyllables. Recent work on disyllables suggests that neither height nor contour alone explain the patterns of AE perception of Mandarin tone. Berkowitz and Strange (in prep) tested native Mandarin speakers and naïve AE speakers on an AX task with Mandarin disyllables. There were 14 stimulus pairs, which were natural productions of the nonsense strings /dima/ and /duma/. The 14 contrasts were chosen so that 10 were predicted to be difficult, 3 were predicted to be of medium difficulty and one was predicted to be easy, based on the overall similarity of their F0 contours. Pairs with small mean F0 differences and similar contours across the two syllables were predicted to be more difficult to discriminate than those with divergent F0 heights or extremely different contours. Five Mandarin subjects and 15 AE subjects participated in this study. As expected, the Mandarin speakers scored at ceiling (individual A' scores ranged from 0.99-1.0), and the AE subjects scored significantly lower, with a group mean A' score of 0.88; individual A' scores ranged from 0.74 to 0.96. The predictions about which contrasts would be most difficult were not clearly supported. The AE subjects' performance could not be predicted solely by the height difference across the two syllables nor by the contour difference between stimuli, as measured in number of turning points. It is by no means obvious that conclusions drawn from studies using monosyllabic stimuli will be supported by results from studies of the

perception of disyllables. There was a wider range of scores for the disyllables that contrasted on initial syllables than those that contrasted on final syllables. Therefore, the study reported here followed up on this by examining only initial contrasts.

1.6 The present study

The present study examined the ability of American English speakers to discriminate Mandarin disyllables. A small control group of Mandarin speakers completed the same experiment; this group was expected to perform with near-perfect accuracy. The stimuli were Mandarin disyllables that contrasted in the initial syllable; there were 21 possible combinations of this sort (see Appendix 32A and Section 2.3). The disyllables were produced naturally, in citation form, so that all the acoustic cues were preserved. All possible initial-syllable contrasts were tested in a name-identity same/different paradigm. The behavioral results were correlated with acoustical measures of amplitude, duration, f0 slope and mean fundamental frequency in order to explore which of these were most critical for discrimination by AE listeners. For example, if mean F0 and offset F0 on a syllable-by-syllable basis are the perceptual cues most important to AE listeners, then disyllables with large mean F0 differences and large differences in offset F0 of the first syllable should be more easily discriminated, while those with smaller differences should be more difficult for AE listeners. Native Mandarin speakers should be able to discriminate all contrasts easily, since they are in their language repertoire.

It was predicted that Mandarin speakers would easily discriminate all Mandarin disyllables in the proposed experimental paradigm, but that AE speakers would perform substantially more poorly due to lack of experience with lexical tone contrasts.

Moreover, since AE listeners were likely to employ a right-hemisphere dominant process to judge the stimuli, as the neurobiological evidence suggests, it was predicted that the two syllables would be treated as one prosodic unit, that is, as a small phrase, rather than as two discrete syllables. Thus, the mean fundamental frequency difference measurement made across the two syllables was predicted to show a stronger correlation with behavioral results than initial syllable-only measurements. It is possible that acoustic differences in the second syllable provided additional information for the AE listeners, causing them to correctly respond to a stimulus pair as different, regardless of the fact that the actual linguistic difference is in the first syllable. More specifically, it was proposed that stimulus pairs with larger differences in average F0 across the two syllables would be easier to discriminate than pairs with smaller differences.

In this study, the stimuli were all of the form /duma/; this was expected to increase the possibility that the subjects attended mainly to the tonal contrasts as instructed, rather than the segments. Rather than testing just one final context as Gottfried and Suiter (1997) did, this study examined all initial tone contrasts in all possible final contexts. In this way, it was possible to test if the second syllable's tone has an effect on the relative difficulty of discrimination regardless of which tone is in the initial position. Because the subjects were naïve, rather than learners of Mandarin, a discrimination task was most appropriate here. In this case, a same/different categorial task with a 1500 millisecond inter-stimulus interval (ISI) was chosen. It has been shown (Werker & Logan, 1985) that naïve listeners must use a phonemic memory strategy to process unfamiliar stimuli at this ISI, rather than just acoustic memory, although this has

not been tested with tones. One previous study of monosyllables with an AX task (Klein, Zatorre, Milner & Zhao, 2001) did not report the ISI used.

1.7 Hypotheses

The following hypotheses were tested:

H1: Mandarin speakers will discriminate the 21 Mandarin initial disyllabic contrasts with near perfect accuracy, but AE speakers will perform significantly more poorly.

Because Mandarin speakers were performing a relatively simple task in their native language, it was predicted that they would score at ceiling on this task. However AE listeners were listening to a foreign contrast, and thus were predicted to perform less well. The relative performance of the AE speakers on particular tone contrasts will be dependent on multiple factors, to wit:

H2: Some tone contrasts in the initial syllable were predicted to be more easily discriminated than others, independent of the following tone. Tone 1 vs. Tone 2 was predicted to be a relatively difficult contrast, because these two tones have a similar F0 offset. Tone 1 vs. Tone 4 was also predicted to be a difficult contrast, because these two tones have similar mean F0. In comparison, Tone 1 vs. Tone 3 was predicted to be the easiest contrast, because of a large difference in mean F0 and a large difference in F0 offset.

H3: The second syllable was predicted to have an effect on the relative ease of discrimination of initial syllable tone contrasts. Specifically, disyllables that end with Tone 3 were predicted to be discriminated better than the other tones. Tone 3 shows the most carryover coarticulation, and therefore, the differences in the F0 pattern in the

second syllable between Tones 13, 23 and 43 are more pronounced than those of the other final tone contexts.

H4: The relative compatibility of the two disyllables was predicted to yield significant differences in discrimination accuracy. Contrasts where both disyllables are compatible or both non-compatible were predicted to be more poorly discriminated because coarticulation would lead to similar slopes in the second syllable for these pairs.

In the second part of the study, correlational analyses were performed to see which acoustic parameters were mostly strongly related to performance on the discrimination task. There were 11 acoustic measures that were derived and entered into correlations with the performance by AE listeners. These were measures of intensity, duration, mean F0, and F0 slope for each syllable separately, onset and offset F0 for the first syllable, and mean F0 for the whole disyllable. Predictions evaluated in this analysis were as follows:

H5: The performance by AE participants was predicted to correlate with the mean F0 of the first syllable and the offset F0 of the first syllable. It was also predicted to correlate with the mean F0 of the disyllable. If AE listeners focused on the contrastive first syllable, it was expected that difference measures of mean F0 and offset F0 for the first syllable would yield the strongest correlation with the measured behavior. However, if the AE listeners focused on the whole disyllable, it was expected that a measure of mean F0 across the two syllables would yield the strongest correlation with the measured behavior. The remaining acoustic measures were not predicted to show as strong correlations with perceptual performance. However, since these measures were known to covary with frequency, it was entirely possible that AE listeners were using them as

secondary cues, and therefore duration and intensity were similarly examined for correlation with the measured perceptual behavior.

Chapter 2. Method

2.1 Participants

AE group: Twenty-seven adult native speakers of American English participated in this experiment. All AE participants were monolingual English speakers who did not consider themselves fluent in any other language. Most subjects had classroom experience in a second language, but none had any exposure to a tone language. Subjects responded to a questionnaire about their language background and musical training, if any. Subjects with conservatory-level musical training were excluded from this experiment, because there is evidence that their musical expertise is brought to bear on these types of experiments, and thus they should be treated as a separate population (Wong, Skoe, Russo, Dees, & Kraus, 2007: Gottfried & Xu, 2008).

MA group: There were 5 participants in the Mandarin-speaking group. These participants were conversant in English, as evidenced by their ability to communicate in English on the phone and via email. These participants grew up with Mandarin as the dominant language in their households, and were using Mandarin on a daily basis in their current living situations.

All participants passed a hearing screening at 25 dBHL at 0.5 kHz, 1 kHz, 2 kHz and 4 kHz bilaterally, (ANSI, 1996) and none reported any speech, language or learning difficulties. All participants were between 18 and 55 years old. They were recruited through online advertisements in New York, posting of flyers, and through personal contacts.

Excluded subjects: There were three additional participants who completed the experiment but whose data were not included in the final analyses. One was a young AE

violinist with extensive musical training. His scores were extremely high, and upon debriefing, he described using his ear training and musical transcription skills to do the task in a highly sophisticated manner. Another was a young AE woman whose scores were extremely erratic, often below chance levels. This indicated that she might not have understood the task correctly. The third was a purported speaker of Mandarin who, upon debriefing, appeared to have some familiarity with Mandarin but by no means a native level. It is likely that her first language was Thai; thus her data were excluded as well.

2.2 Stimuli

The syllable /duma/ with all possible tone combinations was recorded[ii]. This syllable was chosen after consulting with native Mandarin speakers and considering which syllable shapes would be most desirable for acoustic analysis. One combination forms a real word (du2ma3, 'betting on horses') but the remaining disyllables are nonsensical but phonotactially allowed in Mandarin. It was important to choose nonsense words to decrease any lexical influences on performance by native speakers. The syllable combination /duma/ was also a good choice because the Mandarin consonants and vowels are phonetically very similar to English phones produced in the same context, which was expected to make it easier for the American English speakers to focus on the tonal contrasts under study (as opposed to introducing a highly foreign sounding syllable, such as /tcy/). By using a continuant between the two vowels, it was possible to extract F0 data throughout the duration of the two syllables (excluding the initial plosive).

A female native speaker of Mandarin, who is also fluent in English, was recorded in a sound-treated booth with a Shure SM48 microphone placed approximately 5 inches from her lips. The recordings were preamplified with an Earthworks Lab 101 preamplifer and were carefully monitored for peak-clipping, and the levels were adjusted accordingly. The stimuli were digitized with SoundForge 6.0 software as monaural 22.05 kHz, 16-bit files on a Dell PC with a SoundBlaster Pro A/D soundcard. The speaker read all possible tone combinations of /duma/ from a list. The nonsense words were written in both Chinese characters and transliterated in pinyin notation. This list was read and recorded five times. The master recordings were then trimmed into individual .wav files, with minimal silence at the beginning and end of each stimulus, thus ensuring that the eventual stimulus delivery timing was consistent from token to token.

2.3 Acoustic Analysis

The tokens used to create the experimental stimuli were evaluated with Praat software (Boersma, 2001) using a script formerly known as TimeNormalizeF0.praat, and now known as ProsodyPro (Xu, 2007) which measures F0, duration and intensity. Cursors were manually placed at the release of the /d/, the transition between the /u/ and the /m/ as judged by the drop-off of acoustic energy at the beginning of the nasal, and at the final discernable pitch period of the /a/. This separated the stimuli into two syllabic units. For each syllabic unit, measurements of duration and mean intensity were generated. The F0 values for each pitch period were then averaged into ten time-normalized points, along with the relative duration associated with each segment. These raw measures were then used to find mean F0 for each syllable, mean F0 across the disyllable as a whole, slope of the first syllable and slope of the second syllable. Slope

was calculated as a linear regression, that is, the straight line of best fit given the ten frequency and ten time measurements for each syllable. Onset frequency was taken as the first of the ten frequency measures, and offset frequency was the last of the ten.

2.4 Stimulus verification and selection

In order to verify that the stimuli were properly produced and easily identified, two native speakers of Mandarin identified every token by its pinyin tone numbers. Any stimuli that were misidentified by either judge or were found by either judge to be odd-sounding or incorrectly produced was eliminated. After discarding the unusual stimuli, three tokens were chosen for each /duma/ tone combination. If more than three tokens were judged to be correctly produced, they were culled by examining the acoustical measurements taken above. Any tokens that showed extreme differences in intensity, duration, or F0 were eliminated, and then, if more than three tokens remained, they were eliminated with random selection. In this manner, for each /duma/ tone combination, a set of three tokens that were similar but not physically identical was created.

There were 4 possible tones for the initial syllable and 4 possible for the second syllable, which yielded 16 disyllabic combinations. Each of these was paired with all possible disyllables that had the same final tone but a contrasting initial tone. There were three pairings that included tone 33; however, tone 33 is always changed by sandhi rules to tone 23 (tone sandhi refers to systematic changes in tone that are obligatory when certain tone sequences occur) . Therefore these 3 combinations were eliminated. Thus there were 21 tone combinations that contrasted on the initial syllable. There were three tokens for each type, [iii] thus, there were 18 trials (3 of one type x 3 of the other type x 2 orders) for each of the 21 different contrasts for a total of 378 trials whose target answer

was “different.” These 378 trials were distributed into three balanced blocks of 126 trials each, so that the subjects heard each stimulus item an equal number of times in each block.

The trials whose target answer was “same” were prepared in an analogous way; however, the trials where a stimulus item was paired with itself were eliminated. Eliminating the physically identical trials reduced the number of possible trials[iv] from 18 per contrast to 12 per contrast. Thus, there were 252 total same trials, and these were distributed into three balanced blocks of 84 trials. In this way, each block consisted of 126 different trials and 84 same trials, for a total of 378 different trials and 252 same trials. The grand total was 630 trials, in three blocks of 210 trials. The ratio of same pairs to different pairs is 252/378 = 2/3. Alvin experimental control software (Hillenbrand & Gayvert, 2003) randomized each block on delivery to the subjects.

2.5 Procedures

After giving informed consent, the subjects were given a hearing screening. All experimental tasks were delivered via a desktop computer at a comfortable listening level over Sennheiser headphones. An introductory slide show instructed subjects to pay attention to the tone contour of the stimuli, and include monosyllabic Mandarin samples of the four standard tones (see Appendix 1). Subjects were encouraged to listen to the monosyllables twice each. Three blocks of 210 trials were presented, as described above. Stimuli were presented with a 1500 msec inter-stimulus interval, and subjects clicked with a mouse to indicate if the stimuli were the same or different. The experiment was self-paced; the subjects were required to click “go” to proceed to the next stimulus pair

presentation. Subjects were encouraged to take a break after each block. Each block took 30 minutes or less for the American English speakers to complete. Subjects completed a language background questionnaire after completing the experimental task (see Appendix 2). This was collected at the end of the session, because it included questions about musicality, the answers to which could have potentially influenced the experimenter or the subject. The whole experimental protocol was completed in less than 2 hours. Subjects received $20 for their participation. These funds were supplied by the CUNY Graduate Center Doctoral Student Research Grant Program, Competition #4.

Chapter 3. Results: Data reduction

This section reports on the statistical analysis performed on the discrimination task, the acoustic analysis performed on the stimuli, and the correlations between the AE subjects' performance on the discrimination task and a set of acoustic measures drawn from the acoustic analysis.

3.1 Scoring Methods and Statistical Methods for Discrimination Task

Discrimination tasks are frequently scored using methods from Signal Detection theory such as d'. However, A', a non-parametric scoring method, was chosen for use in this study. This was done because there were some individuals who scored with 100% accuracy on some contrasts, which would violate the assumptions of d' but not of A' (Jenkins, 2002; Grier, 1971). A' scores can range from 0.5 (results not different from guessing) to 1.0 (100% accuracy on different pairs and on same pairs). A' scores were computed for each subject's overall performance and for each subject's performance on each of the 21 disyllabic contrasts. Because the A' scores were continuous data, parametric statistics were used to compare AE subjects' A' scores, specifically Analysis of Variance and t-tests. However, MA subjects scored at ceiling, which is a violation of the assumptions for ANOVA. Therefore, when comparing the MA and AE A' scores, a Mann-Whitney U nonparametric test was used. This test is appropriate for continuous data from two independent groups, and can be thought of as the non-parametric version of a t-test (Pallant, 2001).

The data from the AE subjects were also analyzed with regards to a nativelike cutoff. Each of the A' measures was rescored as a proportion, indicating how many of the AE subjects were able to perform at nativelike levels. This is discussed below in Section 3.2.

By converting the participants' performance into nativelike and non-nativelike, the resultant data was dichotomous. Dichotomous data must be evaluated with non-parametric statistics. In this case, Cochran's Q was used for k-related samples, and, if significant, was followed by McNemar change tests for pairwise comparisons (Siegel & Castellan, Jr., 1988). These tests can be thought of as analogous to ANOVA and t-tests. In the case of comparing two independent groups (AE vs. MA) on a dichotomous measure it was most appropriate to use a Fisher's Exact test (Siegel & Castellan, Jr., 1988).

3.2 Results of Perceptual Test

A. H1: Native vs. Non-native overall performance accuracy

The mean A' for the 27 AE subjects was 0.90 (standard deviation 0.08), and individual overall scores ranged from 0.69 to 0.99. The 5 MA subjects had a mean A' of 0.99, and overall individual scores ranged from 0.97 to 1.0 (SD irrelevant due to ceiling effect). Due to non-normal distributions and uneven group sizes, a non-parametric statistical test was used to compare the two groups. A Mann-Whitney U showed that group differences in overall accuracy were statistically significant ($U = 3.00$; $p < .001$). The box and whiskers plot of these data is shown in Figure 2.

The lowest overall perceptual performance by a Mandarin participant was 0.97 and this was used as a cutoff to reexamine the AE participants' overall performance, thus converting the continuous A' data to a dichotomous variable. There were 3 out of 27 AE subjects (11%) who performed at or above a nativelike level. A Fisher's Exact test was computed, comparing the relative proportions of subjects who scored above and below the nativelike cutoff, and was found to be statistically significant ($p = .0003$; a one way test was used). The remainder of the analyses focused on which contrasts were most difficult for AE participants and how that might be related to acoustic measures.

B. H2: Relative difficulty of initial tone contrasts

The second hypothesis stated that some tone contrasts in the initial syllable would be more easily discriminated by AE listeners than others, independent of the following tone context. The A' data from the 21 contrasts were collapsed into six subscores, based on the contrastive initial tone, without regard to the final tone (initial Tone 1 vs. 2, 1 vs., 3…3 vs. 4) and an average A' measure was calculated for each AE participant. The group data for each of the six contrasts are shown in Figure 3 as box and whiskers plots.

The data from the AE listeners was also converted to a dichotomous variable as in the previous section. The lowest mean performance by any Mandarin subject on any of the six initial contrasts was 0.95 A' and so this cutoff was used to rescore the AE subjects' performance. The performance by the AE participants on the six initial tone contrasts as expressed with A' and as nativelike proportions are shown in Table 1.

To test for an effect of initial tone contrast, a repeated measures Analysis of Variance was conducted using A' data. The main effect of initial contrast was found to be statistically significant ($F (5, 22) = 13.45$, $p < .0005$). Because the ANOVA was

significant, a Planned Comparisons analysis was performed. Tone 1 vs. 2 was predicted to be a difficult contrast due to a small difference in F0 offset. Tone 1 vs. 2 was found to be significantly more difficult than any of the other contrasts ($p = .002$ or less) Tone 1 vs. 4 was predicted to be a difficult contrast due to a small difference in mean F0. Tone 1 vs. 4 was significantly more difficult than all contrasts except Tone 2 vs. 3 ($p = .005$ or less). Tone 1 vs. 3 was predicted to be the easiest contrast because of a large difference in both the mean and offset F0. Tone 1 vs. 3 was significantly easier than all other contrasts except Tone 3 vs. 4 ($p = .015$ or less).

The nativelike proportions were analyzed to evaluate this hypothesis as well. Because the nativelike proportions are related samples, that is, the same participants are measured in each proportion, a Cochran Q was performed on these data and found to be statistically significant ($Q = 46.27$, d.f. = 5, $p < .001$). The planned comparisons as explained above were then computed with McNemar change tests, and showed that contrast 1 vs. 2 was significantly more difficult than all other contrasts except 1 vs. 4 ($p = .002$ or less). Contrast 1 vs. 4 was significantly more difficult than all the remaining contrasts ($p = .004$ or less). None of the other two-way comparisons were statistically significant.

C. H3: AE discrimination of final tone contexts without regard to initial contrast

The third hypothesis predicted that the second syllable would have an effect on the relative ease of discrimination of initial syllable tone contrasts. Specifically, disyllables that ended with Tone 3 were predicted to be discriminated better than those that ended with the other tones due to coarticulatory effects. For this analysis, the data were collapsed into four subscores (Final Tone 1, 2, 3 and 4). An average A' measure for

each category was found for each AE participant. As in Section B, these data were also rescored as a proportion of nativelike performance, in this case using the cutoff of 0.96. These data are displayed in Table 2, and a box and whiskers plot is available in Appendix E.

To test for an effect of final context, a repeated measures ANOVA was conducted on the AE's group A' scores. There was a significant main effect of final syllable context ($F(3, 24) = 10.24$, $p = <.0005$). Because the results of the ANOVA were statistically significant, a Planned Comparisons analysis was conducted to test if contrasts preceding Tone 3 were more easily discriminated than contrasts preceding the other 3 tones. This hypothesis was found to be supported as well (Final 1 vs. 3, $t(26) = 3.65$, $p = .001$; Final 2 vs.3, $t(26) = 3.01$, $p = .006$; Final 4 vs. 3, $t(26) = 5.43$, $p < .001$).

The nativelike proportions were also evaluated with Cochran's Q, and found to be statistically significant ($Q = 14.85$, d.f. = 3, $p = .002$). McNemar change tests were conducted on the planned comparisons. Context ma3 was significantly easier than ma4 ($p = .002$). McNemar change tests were computed post-hoc for context ma4, and it was found that ma4 was statistically more difficult than ma2 as well ($p = .031$).

D. H4: AE discrimination of compatible contrasts

The fourth hypothesis predicted that contrasts where both disyllables had the same compatibility would be more difficult to discriminate than those where one disyllable was compatible and one was non-compatible. The slopes of the second syllables were more closely matched when both were compatible or both non-compatible. A repeated measures ANOVA was performed, with the data collapsed into two subscores: one score included contrasts where both disyllables were compatible or both

were non-compatible. The other included contrasts where one disyllable was compatible and the other was non-compatible. This result was statistically significant as well ($F(1, 26) = 29.99$, $p < .0005$). See Appendix F for the box and whisker plot associated with this analysis.

E. Post-hoc analysis of all 21 contrasts for good and poor performers

As described above, some AE subjects performed more accurately than others on the perceptual task. The Mandarin subjects' data was examined to find the lowest A' score that occurred on any of the 21 contrasts by any of the Mandarin-speaking individuals. This was found to be .92, and the AE subjects then had their scores recoded for how many of the 21 individual contrasts they were able to produce at or above .92. These scored ranged from 0/21 to 21/21 and are shown in Figure 4. This distribution appeared bimodal, and it was decided to divide the groups into a "poor" group and a "good" group, with the criterion for "good" defined as native-like performance on at least 12/21 contrasts. Twelve participants (44%) were poor performers and 15 (56%) were good performers by this criterion.

F. Post-hoc analysis of all 21 contrasts for easy and difficult contrasts

After completing the above planned analyses, it was apparent that there was an interaction between the initial contrast and the final context, and that the 21 individual contrasts should be examined individually. The group data for the AE subjects was analyzed in terms of how many of the 27 subjects achieved nativelike performance on each contrast. The A' scores for each of the 21 contrasts by each of the 27 participants ranged from 0.69 to 0.97, whereas the lowest performance on any of the 21 contrasts by a

native Mandarin speaker was 0.92. Therefore, 0.92 was used as a cutoff for nativelike performance as in the previous section. The proportion of AE subjects who performed at or above the nativelike cutoff on a given contrast ranged from .07 (2/27) to .78 (21/27). A Cochran's Q with follow-up McNemar change tests was considered for the analysis of this post-hoc data, on analogy with the previous analyses; however, the number of McNemar change tests required would be too large in relation to the number of data points (210 tests). Instead, each of these nativelike proportions was compared to the binomial distribution. If the 21 contrasts were all equally difficult, it would be expected that they would all evince proportions that were close in value to each other. By this analysis, 5 contrasts were poorly perceived, 6 were easily perceived and 10 were neither poorly nor easily perceived. These data are reported in Appendix D.

In summary, the Mandarin-speaking participants performed with near-perfect accuracy. The lowest overall A' score by a Mandarin-speaking subject was .97 and the lowest score on an individual contrast was .92. American English-speaking participants showed much greater variability in their performance, with the preponderance of participants (24/27) producing overall A' scores below that of the lowest native performer. When only the initial contrast was considered, Tone 1 vs. 2 was most difficult, followed by Tone 1 vs. 4; Tone 1 vs. 3 was easiest to perceive. The difficult contrasts had small mean F0 differences and small offset F0 differences, respectively. When only the final context was considered, disyllables ending in Tone 3 were best perceived. Disyllabic pairs with different compatibility conditions were easier for AE subjects than pairs with matching compatibility conditions. These last two results are

most likely due to the differences in slope of the second syllable. Additionally, when each of the 21 contrasts was considered individually, it was apparent that there were easy, difficult, and middle-level contrasts based on the number of individuals who scored at nativelike levels. By examining each of the 21 contrasts, it was also found that there was a range of performance amongst the AE subjects that fell into better and poorer performance groups. In the next section, the acoustic measures taken were correlated with these behavioral results.

3.3 Correlation of Perceptual Results with Acoustic Analysis

A. Results of acoustic analysis

Each of the 45 tokens used in preparing the trials (15 disyllables x 3 tokens) was analyzed with Praat software (Boersma 2001), using a script called TimeNormalizeF0.praat (Xu, 2007) as described in Chapter 2.3 above. Eleven acoustic measures were either directly available or derived from the output of the script. Each of the measurements taken on each of the three tokens of each type were then averaged together to produce a mean measurement for the disyllable tone combination. After generating these acoustic measures for the 21 individual disyllable types, the two means for each contrast pair were used to create a difference score. These difference scores were used as a measure of (dis)similarity between the two disyllables in a pair; that is, a small difference score implies a small difference between the two disyllable contours. The absolute ranges and difference score ranges are shown in Appendix H.

B. Correlation of acoustics and perceptual performance by AE participants

The difference scores were then correlated with the behavioral scores as measured by A' as described in Section 3.2. Each acoustic difference score was correlated with the perceptual performance of the whole group, and then with the poor performers and the good performers separately. In this way it was possible to examine which acoustic cues correlated most highly with perception, and to see if the two groups' behavior correlated in the same way or a different way. The correlations were evaluated with Pearson r, and a one way test of probability was used. It was appropriate to use a one way test because of the principle that a small difference is harder to perceive than a large difference. The correlations and their associated probabilities are reported in Table 3.

The performance on the perceptual task by the whole AE group was most strongly (and significantly) correlated with the difference scores for mean F0 of the initial syllable (Pearson r = .421; p = .029), offset F0 of the initial syllable (Pearson r = .442; p = .022) and the slope of the second syllable (.445; p = .022). Scatterplots for these correlations are shown in Figure 5. All the other acoustic measures were not significantly correlated with perceptual performance. The correlation coefficients and their associated alpha probabilities are reported in Table 3. The correlations of the subgroups are provided in the table as well; however, they should be interpreted with caution due to the small group size (N = 12 and N = 15 for the poor and good performers respectively). All the correlations were also divided into strong (.4 < r <.69) moderate (.3 < r < .39) weak (.2 < r < .29) and negligible (0 < r < .19) (Cohen, 1977). It can be seen from the table that the poor performers showed weaker correlations than the good performers on 5 measures: mean F0 of /du/ (moderate vs. strong); mean F0 of /ma/ (weak vs. strong); mean F0 of disyllable (weak vs. moderate); slope of /du/ (negligible vs. weak) and duration of /du/

(negligible vs. moderate). Additionally, the poor performers showed a moderate correlation between behavior and the duration of /ma/ while the good performers' correlation was negligible.

These data might be interpreted as implying that the two groups weighted the cues differently. To test this more precisely, the correlation coefficients for the good and poor performers were converted to rank scores. These rank scores were tested with the Friedman two-way analysis of variance by ranks, a non-parametric test of whether two groups are likely to have been drawn from different populations (Siegel & Castellan, Jr., 1988). This test was not statistically significant ($p = .085$) and so it cannot be claimed that the good and poor performers are actually two separate groups based on these samples.

C. Individual contrasts examined with regard to acoustic measures

Each of the 21 contrasts was identified as difficult, easy, or medium in Section 3.2.E above. After the acoustic analysis was conducted, each difficult and each easy contrast was reexamined with regards to the acoustic measures as reported in Appendix H. The easy contrasts generally had larger first syllable mean F0 differences and first syllable offset F0 differences, and the difficult contrasts generally had small first syllable mean F0 differences and small first syllable offset F0 differences. Two of the easy contrasts were noted to have particularly large difference scores for duration and intensity of /ma/: Tone 13 vs. 43 and Tone 23 vs. 43. So it is possible that in individual cases, there were acoustic cues other than F0 values that affected perception, even though these cues were not correlated with overall behavior.

In summary, the performance by the AE speakers on the perception of individual Mandarin disyllabic tone contrasts showed strong correlations with three parameters: the mean F0 difference of the first syllable, mean F0 offset difference of the first syllable, and difference in slope of the second syllable. When the group was split into good and poor performers, the picture changed somewhat, but the groups were small, so it is important to be cautious in interpreting these data. In general, whereas the poor performers' scores were associated most strongly with the same acoustic cues as for the good performers, the strength of the associations were weaker.

Chapter 4. Discussion

This study examined perception of disyllabic Mandarin tones by naïve American English listeners. There are 105 possible disyllabic tone contrasts in Mandarin[v]. Here, all 21 possible disyllabic initial tone contrasts with the second syllable held constant were tested in an AX task. Initial contrasts were chosen for this study based on a pilot study (Berkowitz & Strange, in prep.) that demonstrated that responses to the initial contrasts showed a wider range of accuracy than responses to the final contrasts. This was the first such study that examined the performance of naïve AE listeners on all initial tone contrasts with disyllables in an AX task. In addition, acoustic data were reported for disyllables produced by a female native Mandarin speaker, and these data were then correlated with AE listeners' behavior. Some acoustic data drawn from male speakers' disyllabic productions were previously reported by Xu (1994); however, this was the first study (to the researcher's knowledge) to report acoustic data on disyllables produced by a female and the first to examine the correlation between acoustics and discrimination by nonnative speakers.

Studies of perception of monosyllabic stimuli have been used to draw various conclusions concerning which acoustic cues are used by non-native listeners. These studies suggest that F0 height is the most important cue for nonnative perception. Examination of the perception of disyllables, however, exponentially increases the possible explanations. In this study, there were multiple factors that were shown to have an effect on the relative difficulty of a disyllabic tone contrast: the particular initial tone contrast, compatibility status of the two stimuli, and the identity of the final syllable. Each of these factors had a significant effect on the relative difficulty of the tones. For

each of these factors, there were related acoustic measures that were shown to correlate with the behavior seen. Based on these outcomes, it becomes clear that conclusions that have been drawn about cross-language perception of monosyllabic tone do not necessarily apply to perception of disyllabic tones.

4.1 Native vs. non-native performance overall

It was expected that Mandarin speakers would perform the task with near perfect accuracy, and this was found to be so. These were naturally produced stimuli, and, while they were lexically nonsensical, they were well formed. AE listeners were expected to perform less well, due to inexperience with lexical tone contrasts, and this also was confirmed. Thus, it can be concluded that language experience influences the ability to discriminate disyllabic tones and that performance on the task is language-based.

4.2 Initial contrasts: effect of mean F0 and offset F0

The stimuli were chosen so that they contrasted on the initial syllable and so it was predicted that syllable-initial contrasts would have an effect on discrimination. The tone contrasts that were difficult for the AE listeners were the ones that had small F0 average differences or small F0 offset differences in the first syllable. In contrast, F0 onset differences were not significantly correlated with relative discrimination difficulty, even though the range of difference scores for onsets and offsets were similar. Tone contrasts with large overall mean f0 differences included Tone 1 vs. 3 (94 Hz) and Tone 3 vs.4 (102 Hz). Performance was best on these contrasts with large mean F0 differences. It is not obvious why an onset difference would be discounted by the

listeners; perhaps an initial rise is interpreted as an unimportant stylistic cue rather than a meaningful cue.

4.3 Final contexts: effect of compatibility and slope

When the subjects' responses were grouped by final syllable context, it was seen that those stimuli that ended in Tone 3 were more easily discriminated than those that ended in Tones 1, 2 or 4. Final syllable context had an effect because there was extra information available in syllables that ended in Tone 3. This is because these stimuli had the most surface coarticulation. For this speaker, the three disyllables that end in Tone 3 are maximally differentiated by differences in the second syllable's slope, duration and intensity. The AE subjects were able to rely on these surface distinctions to differentiate the stimuli. Thus, the coarticulation that occurred on the second syllable introduced additional acoustic information for the AE listeners, even though the tones were tonologically the same. Coarticulatory patterns are learned by the native speaker, but a naïve listener does not know the corpus of rules that are in operation. It was hypothesized that the naïve AE listeners are accustomed to listening to F0 across a phrase, and not syllable-by-syllable, so in cases where the coarticulation has a differential effect on the final syllables, it served to make the task easier by making the stimuli more distinct. The relative compatibility of the two tones in the disyllables introduced changes to the F0 contour; specifically, the offset F0 of two disyllables with the same final context will be similar if both are compatible or both are non-compatible. If the two disyllables are of opposite compatibility, the offset of the first syllables will be farther apart in value. The statistically significant correlation of the slope of the second syllable with the perceptual performance by the AE listeners supports the idea that they are attending to

the second syllable and not only the first. However, the pattern of performance suggests that they are attending to the F0 slope more than to the F0 mean.

One outcome that was not predicted was that the Tone 4 context was the most difficult of the four contexts (only 2/27 AE subjects performed at nativelike levels). In this case, the falling F0 of the second syllable influenced the AE listeners to judge that the disyllables were the same, even though the first syllables were different. It can be argued that in this case, the typical English falling prosodic contour is a stronger cue than the differences in F0 occurring in the first syllable. This unexpected result also supports the proposition that the AE subjects listen to the disyllables as a unit; or, if listening on a syllable-by-syllable basis, they are unduly influenced by the final fall, disregarding the differences in the initial syllables.

4.4 Height vs. contour

Previous work on monosyllables argued that AE listeners rely primarily on F0 height rather than f0 contour (Chandrasekaran, Krishnan & Gandour, 2007; Chandrasekaran, Gandour & Krishnan, 2007; Gandour, et al., 2000). The results of the present study show that while F0 height differences are important in the first syllable, slope is what is important in the second syllable. Thus, the height/contour dichotomy so often presented for monosyllables is an oversimplification. American English listeners are attending to the slope of the second syllable.

There have been multiple perceptual studies of monosyllabic tones that found AE listeners were most likely to confuse Tones 2 and 3 in an identification task (Gottfried & Suiter, 1997; Wang, Spence, Jongman & Sereno, 1999) and to make errors of discrimination between Tones 2 and 3 (Gottfried & Ouyang, 2005). These monosyllables

were seen as the most difficult pairing for AE listeners because they are both contour tones that begin and terminate at similar F0 values. However, upon moving to disyllables, all tones except Tone 11 become contour tones. Tone 2 and Tone 3 do have small mean F0 differences and small offset F0 differences in monosyllables, but in disyllables, the offset differences are at least 50 Hz. In this study, when the data were collapsed by the six initial tone contrasts (Table 1), Tone 2 vs. 3 fell into the middle difficulty group of disyllables (10/27 nativelike). In contrast, only 1/27 performed at a nativelike level on Tone 1 vs. 2, which is typically considered an easy contrast in monosyllables. These results show that conclusions drawn from research with monosyllables cannot be generalized to perception of lexical tone contrasts in continuous speech utterances. By moving to more ecologically valid stimuli, although still lab-recorded speech, and still only disyllabic, the results discussed here are quite disparate from results reported for monosyllabic tones.

4.5 Attending to one or two syllables

One hypothesis (H5) proposed that AE listeners would attend to the two syllables together when deciding if a dyad was the same or different. Although the mean F0 across the two syllables was moderately correlated with behavior (statistically significantly only for the good performers) there is certainly evidence that the subjects used information from the second syllable. The slope of the second syllable correlated as strongly with performance accuracy as the mean F0 of the first syllable and the offset F0 of the first syllable. So the subjects did make use of acoustic information in both syllables, just not the parameter that was originally predicted.

Previous work has shown that Tone 4 in non-final position was particularly difficult to identify for naïve AE listeners (Broselow, Hurtig & Ringen, 1987). However, in this study, the contrasts that had Tone 4 in initial position were generally well discriminated, except for Tone 12 vs. 42 and Tone 14 vs. 44. Thus it would seem that the falling F0 of Tone 4 was not intrinsically more difficult. Rather, it is the small mean difference in F0 in these contrasts that create the difficulty. This could be tested more directly in future work by moving to longer stimuli or by using a similar study to the one described here, but with all possible Tone 4 combinations in both positions. Conversely, when disyllables ended with Tone 4, subjects tended to overlook the differences in the first syllable and label the two stimuli as being the same. Contrasts in the /ma4/ context were the most poorly perceived. It can be argued that the subjects in this case treated the stimuli as mini-phrases, and the final drop was similar enough to English sentence-final prosody that the subjects were unduly influenced by the final Tone 4 context.

4.6 Limitations

When planning future work, it is important to remember the lexical advantage that native speakers have in these studies. Listening to a string of meaningless syllables, as the AE listeners did, is quite different from listening to a nonsensical string of real words, such as “spoon horse.” By filtering the stimuli so that the consonants were obscured, the comparison between the two groups would be more equivalent. However, filtered disyllables can sometimes sound monosyllabic (Wong, 2008) so this methodological approach will need further inquiry. A filtering method that would remove enough of the signal to obscure the identity of the consonants, but not so much as to obscure the syllable boundary, would be ideal.

4.7 Future studies

Future directions in response to this dissertation include using the same or a similar experimental paradigm with child participants, particularly speakers of Mandarin, speakers of AE, and speakers of AE who were adopted from Mandarin-speaking orphanages. The internationally adopted children are of particular interest due to their early exposure to a tone language that was then discontinued. This raises many interesting theoretical questions about attrition and reactivation of early language experience. Similarly, this paradigm can be easily used with naïve subjects with stimuli from other tone languages. There are no studies of cross-language perception of disyllabic stimuli in other tone languages to this researcher's knowledge. There is one study in progress which will examine ERP responses to disyllabic Mandarin tones in native and naïve listeners (Yu, in prep).

Performance by the AE subjects tended to be bimodally distributed into good and poor overall performance. The difference between these two groups could have been one of quantity or quality. That is to say, it is possible that the good performers were just more attentive and better at the task, or it is possible that they were applying a different skill set to the task. The correlations of performance of the two groups with acoustic measures were computed, and the results were somewhat different for the two groups. Primarily, the correlations with two of the major acoustic cues (mean F0 of the first syllable and slope of the second syllable) were weaker for the poor group than for the good group, whereas F0 offset of the first syllable correlated more strongly with poor performers' scores than for good performers' scores. However, with the small number of subjects in each group, it is important not to overinterpret these differences. Indeed, the

rankings of the acoustic measures by group were not statistically significantly different overall for the two groups, leading to the conclusion that although it appears that there are two groups due to the bimodal appearance of the data, it may be that individual perceptual differences were actually on a continuum. This could be explored by testing more participants.

In the future, a more difficult task might reveal more individual differences in naïve perceivers' performance by spreading the range of responses. A move to sentence level stimuli or to an ABX task would add an additional memory load, and this might serve to differentiate performers even further. In the same vein, adding an auditory distracter between the stimuli would increase the memory load and might serve to spread out the range of performances even more. The AX task was used here because of its relatively low memory load, and also with an eye to a future studies with children who have shorter memory spans.

The data collected from the language background questionnaires were examined for any obvious differences between the strong and weak performers. No differences in languages studied, musical experience, or duration or intensity of said pursuits were apparent. It would be interesting to actually test the subjects' musical ability rather than just question them about it. Similarly, the debriefing responses were intriguing. Many people included comments about listening to the second syllable, and some participants mentioned the "strength" or "length" of one syllable vs. the other. These impressions could be collected in future studies with a more specific questionnaire rather than informal, open-ended discussion, and then correlated with performance as well.

When examining these results in light of two frequently cited models of cross language speech perception, the Perceptual Assimilation Model (Best, 1995) and the Speech Learning Model (Flege, 1995), it is unclear how to proceed. Only a subset of the possible disyllabic tone combinations has been attempted, and so it might be premature to try to fit these data into one theoretical framework or another. However, these data do serve to point out that both these models are based on comparing sets of phonemes only. In the case of comparing fundamental frequency contours from one language to another, these models do not really offer the needed framework. As more data are collected, it may become clearer how to fit cross language prosodic studies into the body of literature and the developmental and learning models.

Tone languages are found on five continents (Yip, 2002) and Mandarin is the most widely spoken language on the planet. The study of the perception of tone languages by non-tone speakers has focused on monosyllables in the past. Here it has been argued that stimuli must become more complex to properly reveal the underlying perceptual behaviors shown by nonnative listeners. AE subjects relied on F0 mean height and offset height in the first syllable, but on contour (as measured by slope) in the second syllable. This paper examined initial tone contrasts only, that is, 21 out of a possible 105 disyllabic contrasts, and as such can be seen as a first step toward more natural stimuli and thus toward a fuller explanation of cross-language tone perception.

Chapter 5. Tables

Table 1. Performance by AE subjects on the six initial tone contrasts without regard to the final context. Scores are reported as A' and as the proportion of subjects (out of 27) who performed at or above the native cutoff.

Contrast	Mean A' (SD)	Proportion nativelike
1 vs. 2	0.85 (0.08)	1 / 27
1 vs. 3	0.93 (0.08)	15 / 27
1 vs. 4	0.89 (0.08)	2 / 27
2 vs. 3	0.89 (0.10)	10 / 27
2 vs. 4	0.91 (0.08)	13 / 27
3 vs. 4	0.92 (0.08)	13 / 27

Table 2. Performance by AE subjects on the four final tone contexts without regard to the initial contrast. Scores are reported as A' and as the proportion of subjects (out of 27) who performed at or above the native cutoff.

Context	**Mean A' (SD)**	**Proportion nativelike**
Ma 1	0.89 (0.09)	7 / 27
Ma 2	0.90 (0.07)	8 / 27
Ma 3	0.92 (0.08)	12 / 27
Ma 4	0.88 (0.08)	2 / 27

Table 3. Strength of correlations of all AE participants' perceptual performance, and then subdivided into good and poor performing groups. Perceptual performance is correlated with multiple acoustic measures. The three acoustic measures whose correlations with all participants' behavior were statistically significant are listed first.

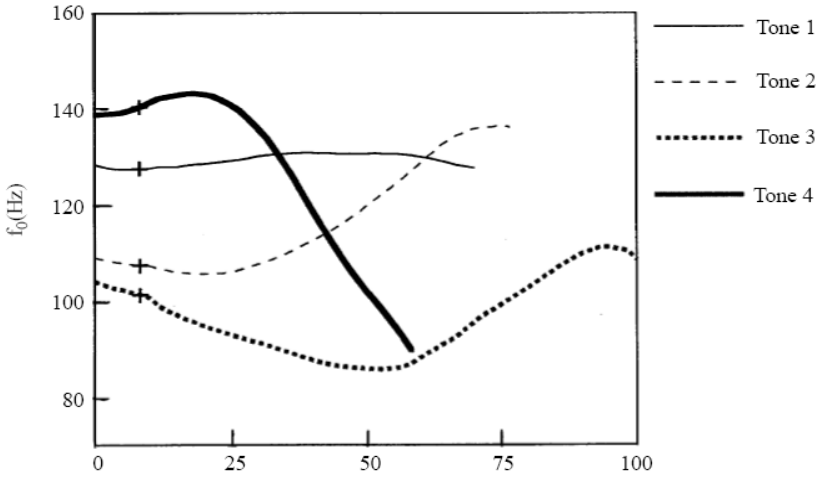
Difference score On Acoustic Measures	**All AE participants N = 27 Pearson r (p value)**	**Poor performers N = 12 Pearson r (p value)**	**Good performers N = 15 Pearson r (p value)**
Mean F0 of /du/	**.421 (.029)***	.360 (.054)	**.439 (.023)***
Mean F0 of offset of /du/	**.442 (.022)***	**.432 (.025)***	**.408 (.033)***
Mean Slope of /ma/	**.445 (.022)***	**.387 (041)***	**.454 (.019)***
Mean F0 of disyllable	.346 (.062)	.276 (.113)	**.378 (.046)***
Mean F0 of /ma/	.329 (.073)	.203 (.188)	**.407 (.034)***
F0 onset of /du/	.344 (.063)	.336 (.068)	.321 (.078)
Slope /du/	.200 (.192)	.124 (.296)	.245 (.142)
Duration /du/	.268 (.120)	.171 (.229)	.327 (.074)
Duration /ma/	.204 (.187)	.357 (.056)	.045 (.423)
Intensity /du/	-.051 (.413)	.073 (.377)	-.156 (.250)
Intensity /ma/	.366 (.052)	.332 (.071)	.355 (.057)

Table 4. Acoustic measures used in correlations. Absolute range and range of difference scores are given.

Acoustic measure	minimum	maximum	Range of difference score
Mean F0 of /du/	177 Hz	244 Hz	4.80 – 109.17 = 104.4 Hz
Mean F0 of offset of /du/	184 Hz	296 Hz	1.18 – 101.00 = 99.82 Hz
Mean Slope of /ma/	-358	166	8.00 – 188.00 = 180.00
Mean F0 of disyllable	195 Hz	284 Hz	3.14 – 72.16 = 68.96 Hz
Mean F0 of /ma/	180 Hz	278 Hz	4.54 – 57.82 = 53.26 Hz
F0 onset of /du/	213 Hz	343 Hz	4.04 – 130.57 = 126.50 Hz
Slope /du/	-546	271	111.00 – 817.00 = 706.00
Duration /du/	177 ms	244 ms	1.75 – 59.02 = 57.25 ms
Duration /ma/	285 ms	501 ms	0.42 – 216.40 = 215.6 ms
Intensity /du/	79 dB	82 dB	0.02 – 2.19 = 2.17 dB
Intensity /ma/	74 dB	83 dB	0.00 – 4.70 = 4.70 dB

Chapter 6. Figures

A



B

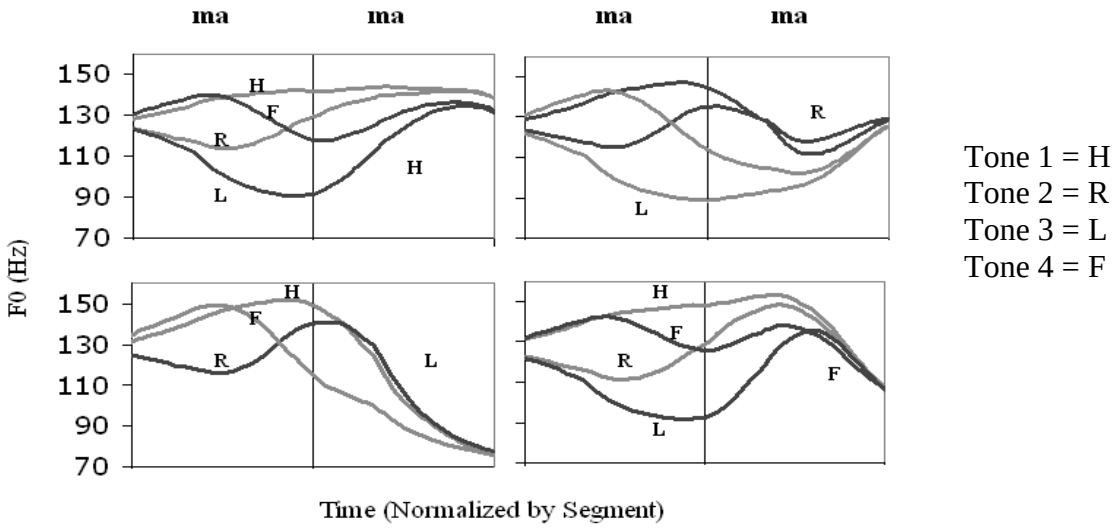


Figure 1. Tracings of average fundamental frequencies for single syllables (A) and disyllables (B). The disyllable tracings have been time-normalized. Adapted from Xu, 1997.

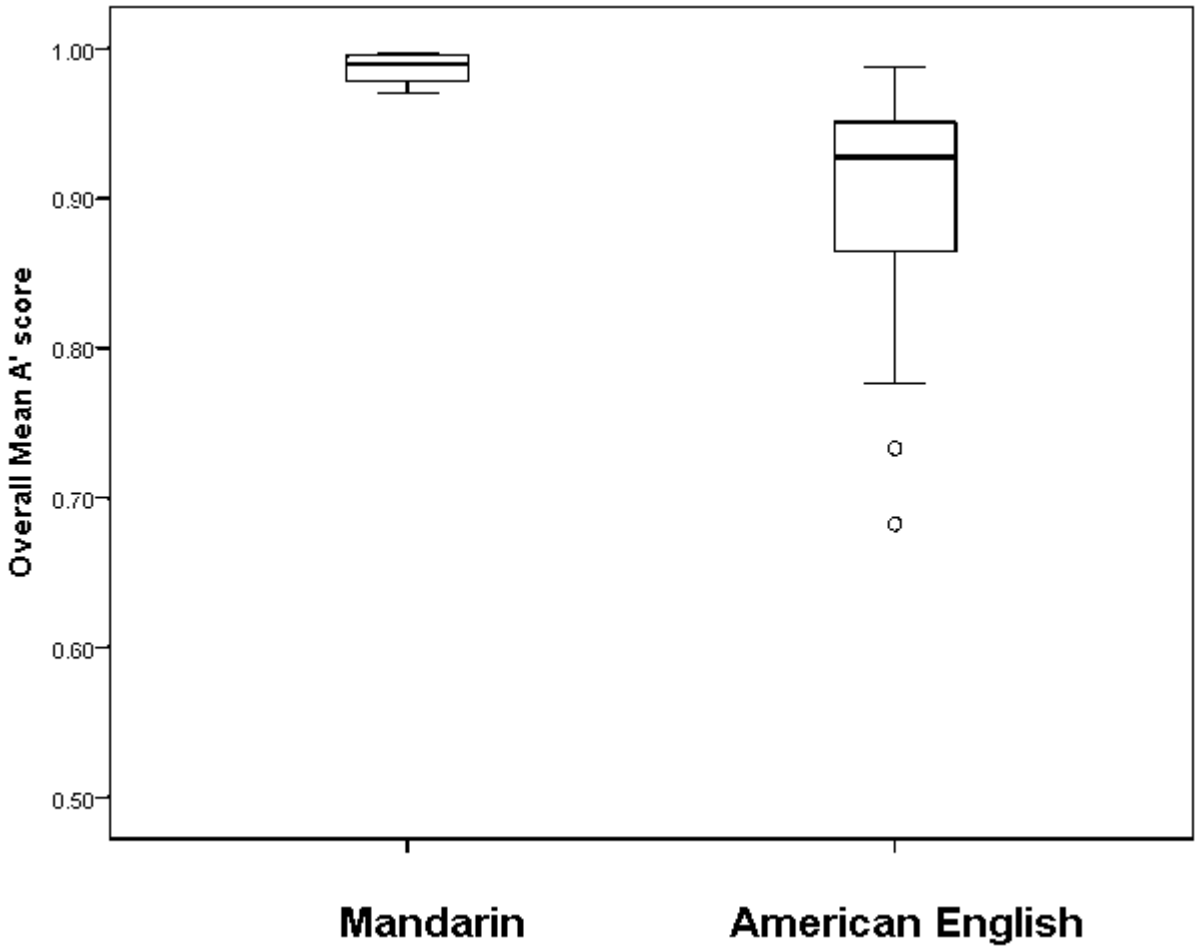


Figure 2. Comparison of overall performance on the perceptual task by Mandarin listeners (N = 5) and American English listeners (N = 27). Performance is reported as an A' score.

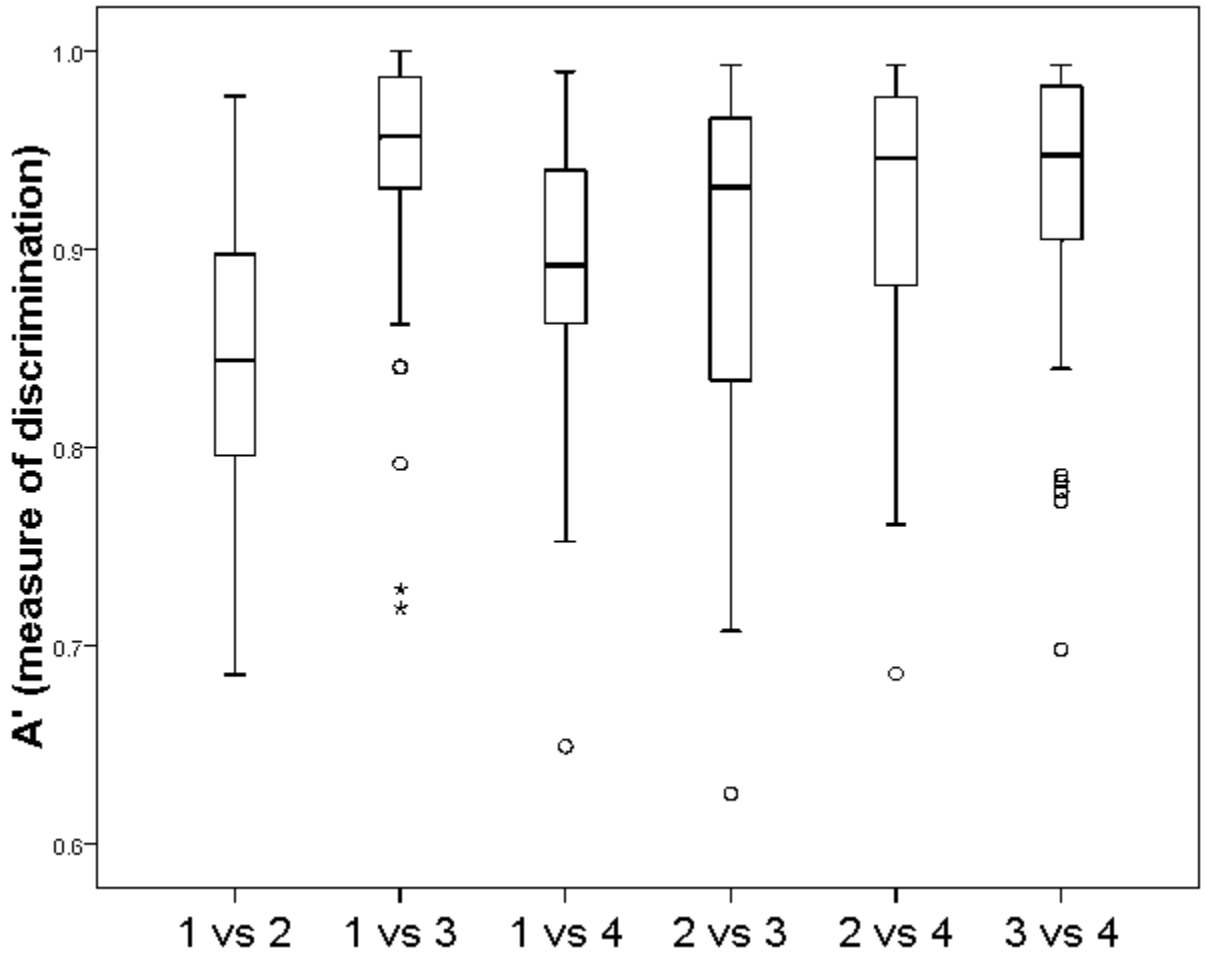


Figure 3. Performance by AE subjects on the six initial tone contrasts without regard to the final context.

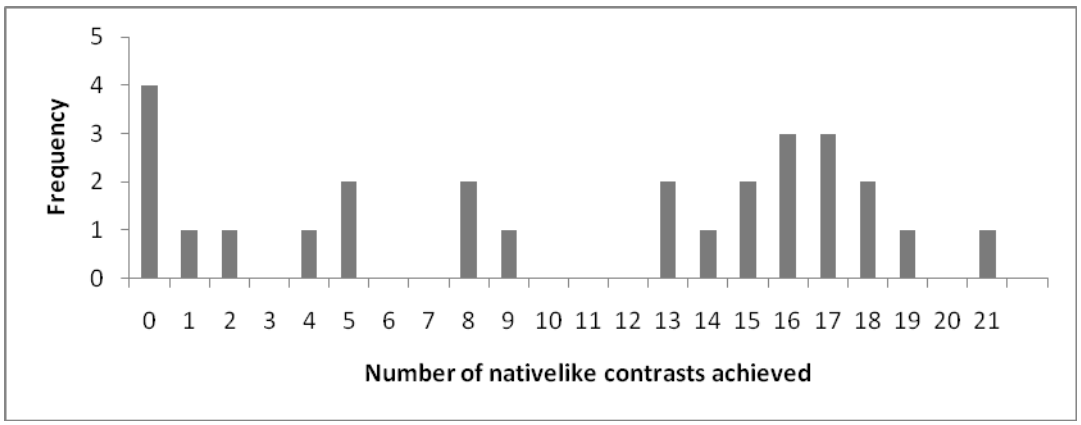


Figure 4. Histogram showing participants' performance as the number of contrasts upon which the subject could perform at nativelike levels. One subject was able to perform at or above the native cutoff for all 21/21 contrasts; 4 subjects were not able to achieve nativelike performance on any contrasts.

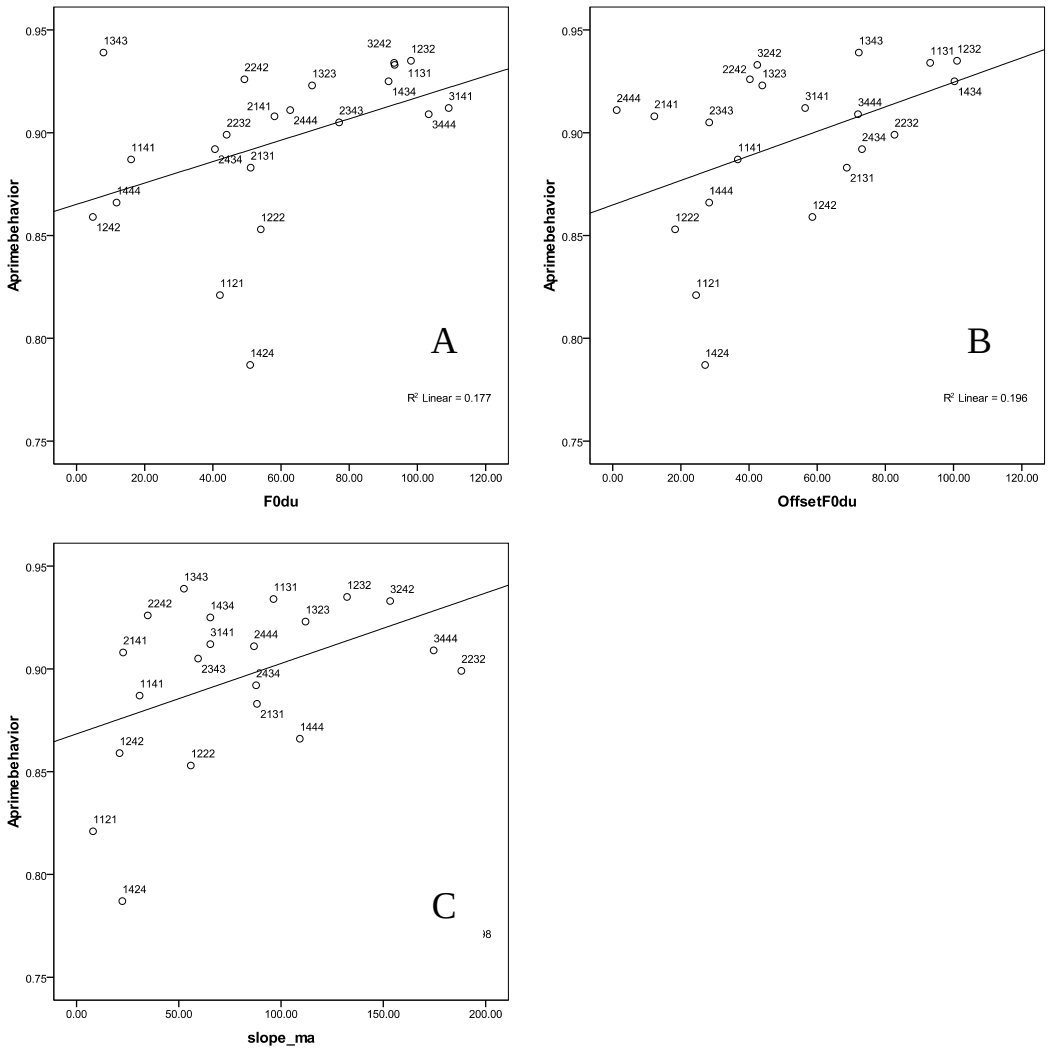


Figure 5.

A: Correlation of perceptual performance by AE participants and difference scores of the mean fundamental frequency of the first (contrastive) syllable.

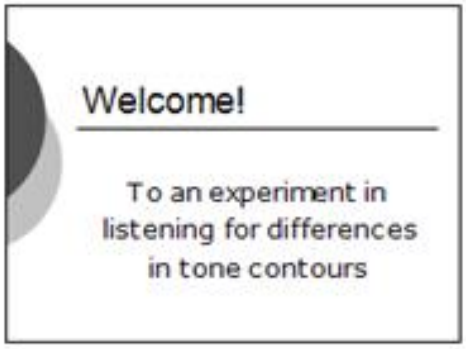
B: Correlation of perceptual performance by AE participants and difference scores of the offset fundamental frequency of the first (contrastive) syllable.

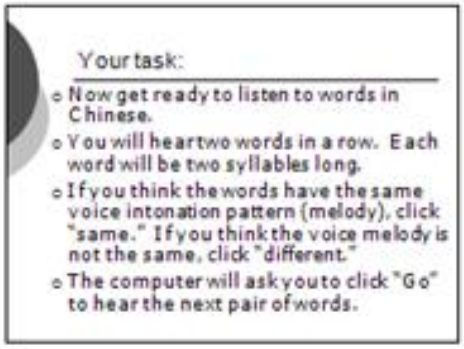
C: Correlation of perceptual performance by AE participants and difference scores of the slope of the fundamental frequency of the second (non-contrastive) syllable.

Label “1424” represents the contrast pair of du1ma4 versus du2ma4.

Chapter 7. Appendices

Appendix A: Instructions to participants





Appendix B: Language Background Questionnaire

PAGE 1 of 3

Language Background Questionnaire

Name of Experiment: DISS [redacted] - SB - ADULTS -
Date: _____
Name: _____ Date of Birth: _____ Gender: _____
Address: _____
Telephone Numbers: (Home) _____ (Work) _____
Birthplace: _____ _____
Town/City State/Country
Father's Birthplace: _____
Languages your father speaks fluently: _____
Mother's Birthplace: _____
Languages your mother speaks fluently: _____
Places in which you have lived for more than 1 year:

City/State/Country	Years
_____	from _____ to _____
_____	from _____ to _____
_____	from _____ to _____
_____	from _____ to _____

If you have lived in more places please check here _____ and continue on the back.

As a child, what languages were spoken in your home? (for example, by parents, guardians, grandparents, or relatives) _____

What languages do you speak fluently and understand without effort?
1. _____ 2. _____ 3. _____
What language do you consider your "mother tongue"? _____
What language(s) did you speak/understand as a child (before going to school)?
1. _____ 2. _____ 3. _____
What language(s) were used in your classrooms in elementary school?
1. _____ 2. _____

What language(s) did you study as a foreign language in school?
1. _____ (Circle all applicable: elementary, junior high, [redacted] high, college).
Number of semesters _____. Did you have a native speaker of the language as a teacher or tutor? No____ Yes____ (number of semesters with native speaker____)
Rate your fluency and understanding of this language by checking one of the following which best describes your mastery of this language:
   a. speak/understand like a native speaker_____
   b. speak with a mild accent and understand native speakers with little or no difficulty_____
   c. speak with an accent and understand, but with some effort_____
   d. speak and understand, but with effort_____
   e. cannot speak or understand this language at all_____

Foreign Language Study (continued) IF YOU STUDIED ADDITIONAL LANGUAGES:

2. _____ (Circle all applicable: elementary, junior high, [redacted] high, college). Number of semesters _____. Did you have a native speaker of the language as a teacher or tutor? No____ Yes____ (number of semesters with native speaker____)
Rate your fluency and understanding of this language by checking one of the following which best describes your mastery of this language:
   a. speak/understand like a native speaker_____
   b. speak with a mild accent and understand native speakers with little or no difficulty_____
   c. speak with an accent and understand, but with some effort_____
   d. speak and understand, but with considerable effort_____
   e. cannot speak or understand this language at all_____

3. _____ (Circle all applicable: elementary, junior high, [redacted] high, college). Number of semesters _____. Did you have a native speaker of the language as a teacher or tutor? No____ Yes____ (number of semesters with native speaker____)
Rate your fluency and understanding of this language by checking one of the following which best describes your mastery of this language:
   a. speak/understand like a native speaker_____
   b. speak with a mild accent and understand native speakers with little or no difficulty_____
   c. speak with an accent and understand, but with some effort_____
   d. speak and understand, but with considerable effort_____
   e. cannot speak or understand this language at all_____

Have you ever studied Phonetics (the scientific study of speech sounds) in high school or college level in a linguistics, speech science, or foreign language class? YES / NO
If YES, have you ever done phonetic transcription? YES / NO
If YES, how much? _____

Do you have normal hearing? YES / NO
Did you at any time have therapy for a speech or language problem?_____

What do you consider your racial/ethnic background to be? Check all that apply.
(Optional: you need not answer)

Caucasian _____ Native American _____
African American _____ Pacific Islander _____
Hispanic _____ Asian American _____
Other- please specify _____

Language Background Questionnaire, Continued

Musical training:

_____ I have never studied music or voice. (stop here—you're done!)

_____ I have studied music or voice

Name of instrument #1:_____

Age you started _____

Age you stopped _____

Level of training (private lessons, school orchestera, etc)

_____

Name of instrument #2:_____

Age you started _____

Age you stopped _____

Level of training (private lessons, school orchestera, etc)

_____

Name of instrument #3:_____

Age you started _____

Age you stopped _____

Level of training (private lessons, school orchestera, etc)

_____

Appendix C

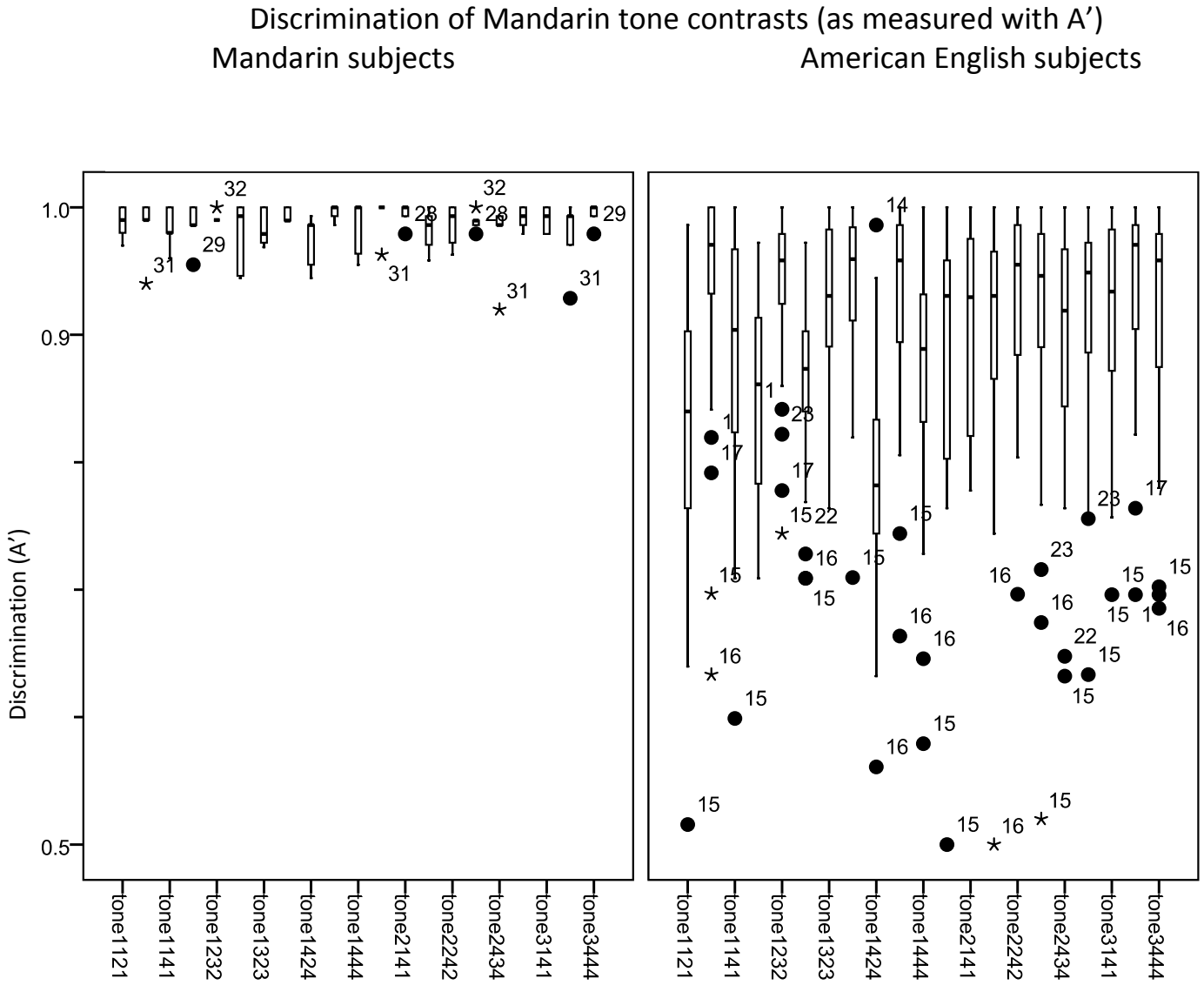


Figure C1: Performance on a same/different task by MA (N = 5) and AE (N = 27) subjects.

Appendix D

Table D1: AE performance on all 21 contrast pairs, rescored as proportion of participants (out of 27) who performed above the native cutoff. Those that are different from chance (as measured with binomial test) are indicated with # for those contrasts that few participants perceived and with & for those contrasts that many participants perceived.

Syllabic Context				
Tone Contrast	ma1	ma2	ma3	ma4
1 vs. 2	**5 #**	**6 #**	15	**2 #**
1 vs. 3	**21 &**	**20 &**		**19 &**
1 vs. 4	12	**3 #**	**20 &**	**7 #**
2 vs. 3	14	16		12
2 vs. 4	14	18	**19 &**	15
3 vs. 4	15	**19 &**		15

Appendix E

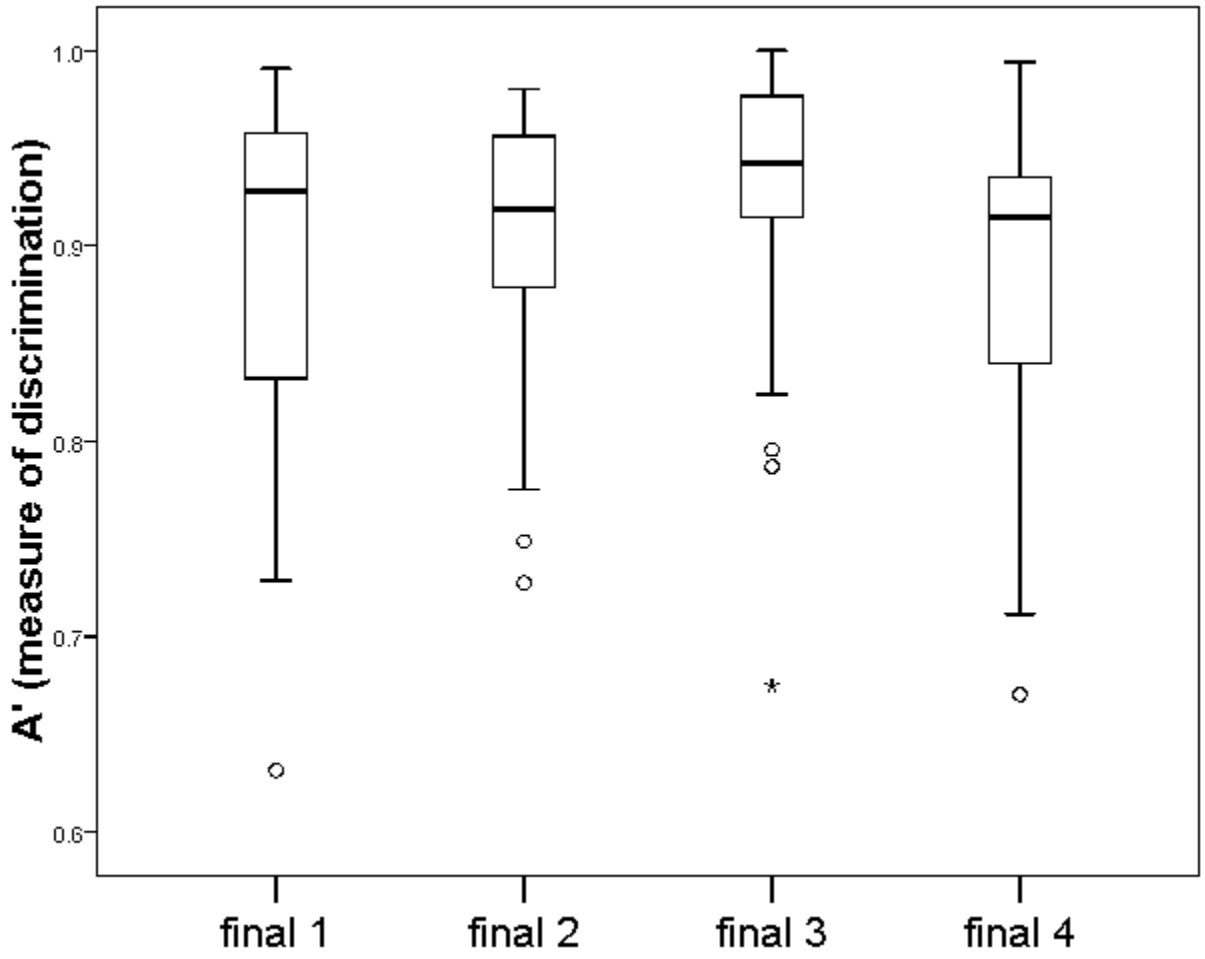


Figure E1: Performance by AE listeners on final tone contexts. Subjects performed best in the context of Tone 3.

Appendix F

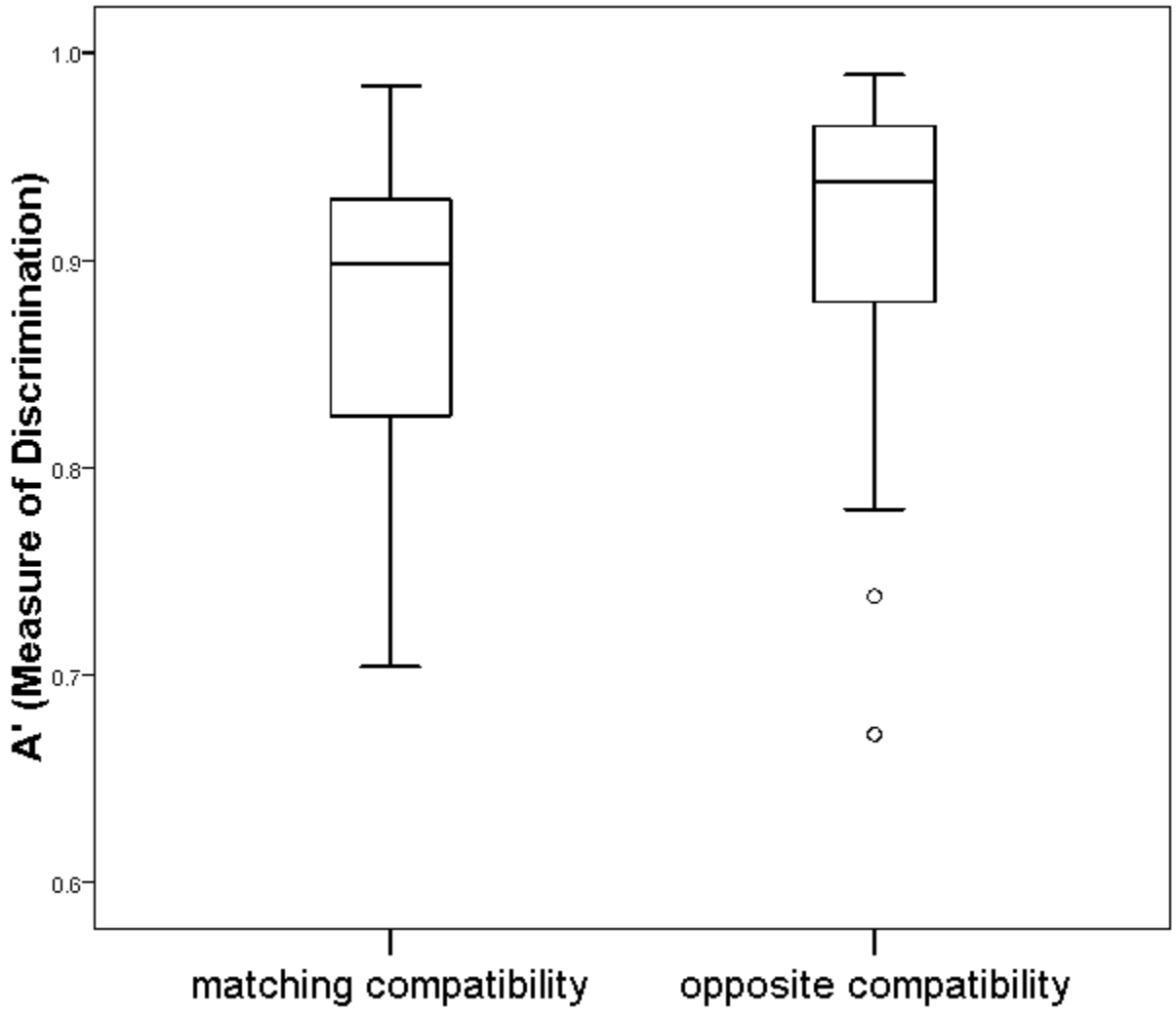


Figure F1. Performance by AE subjects on contrast pairs that have matching compatibility (both compatible or both non-compatible) and non-matching compatibility (one compatible and one non-compatible).

Appendix G

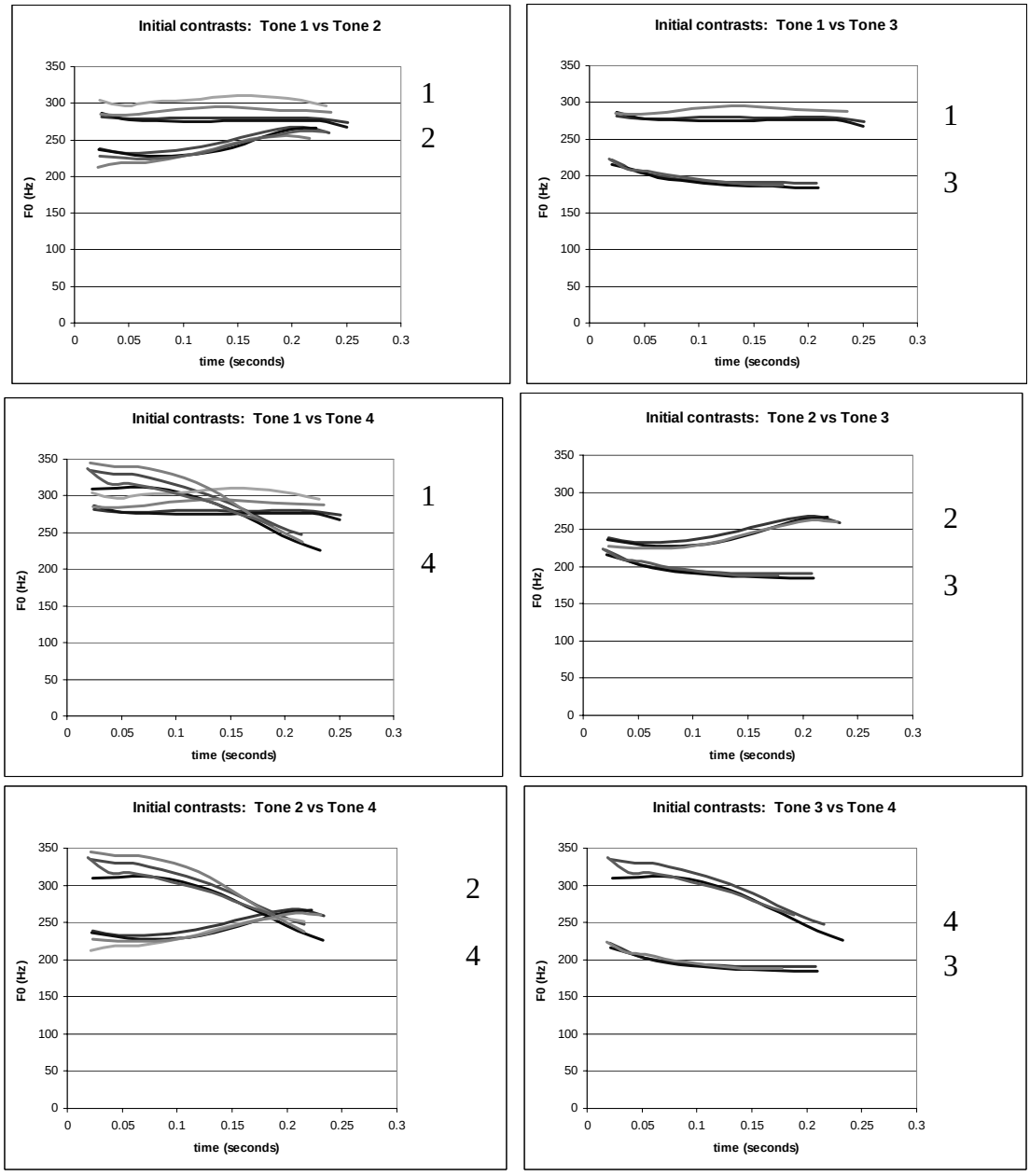


Figure G1. Fundamental frequency tracings of the initial syllable of the disyllabic contrasts without regard to the final context. Each curve represents the average of the three tokens used as stimuli.

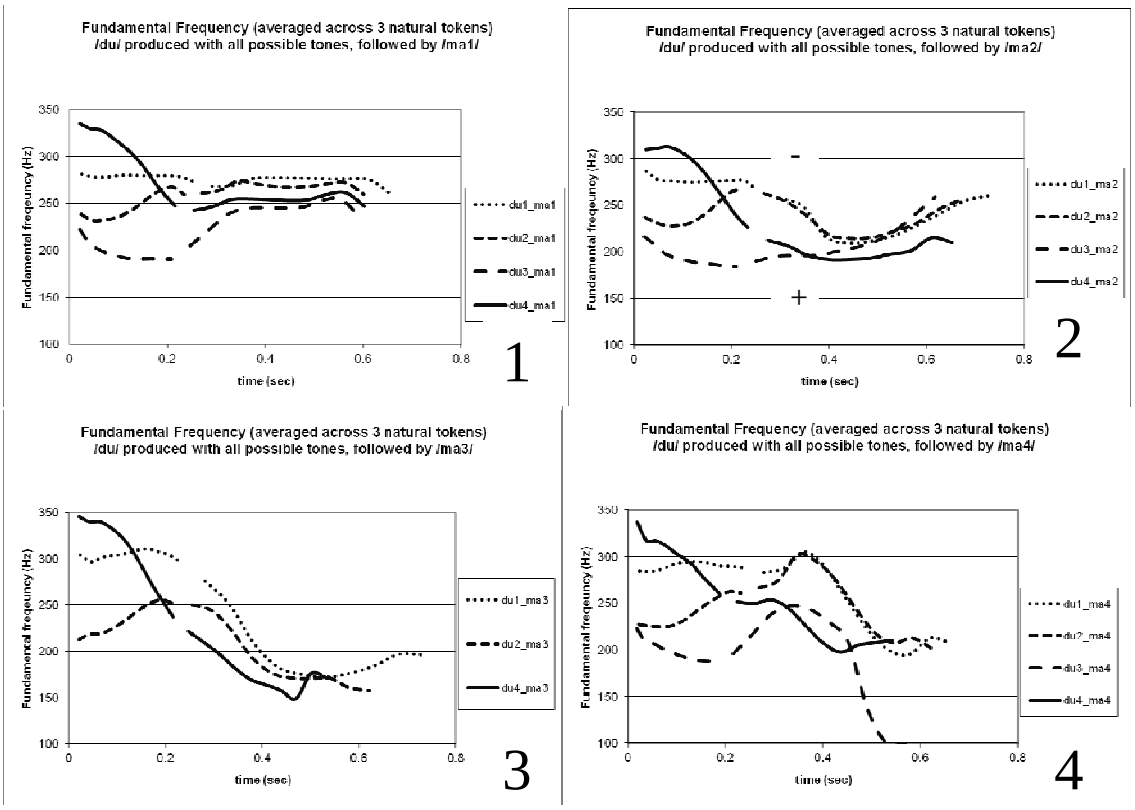


Figure G2. The fundamental frequency contours grouped by final context. Each curve is the mean of the three tokens used in the stimulus set. In panel 1, all the stimuli have Tone 1 in the final position; note that when preceded by Tone 3, Tone 1 is not flat and not high. In panel 2, the two compatible (Tone 32 and Tone 42) and the two non-compatible (Tone 12 and Tone 22) have been marked with "+" and "-" respectively. Pairs of tones with opposite compatibility have been shown to be more easily discriminated. In Panel 3, note that there is a large difference in duration and height amongst the three stimuli.

Appendix H

Acoustic difference scores for each of the 21 tone comparisons.

Acoustic Difference scores for each tone comparison

tone comparison		**mean F0 /du/**	**F0 offset**	**slope /ma/**	mean F0 /ma/	mean F0 /duma/	Duration /du/	duration /ma/	intensity /du/	intensity /ma/	F0 onset	slope /du/
1131	easy	**93.16**	**93.15**	**96.30**	39.34	66.15	21.77	23.89	0.12	2.23	68.87	119.60
1232	easy	**98.11**	**101.00**	**132.25**	14.61	57.45	14.74	62.00	1.70	1.41	83.16	110.74
1343	easy	**7.92**	**72.21**	**52.53**	21.80	10.35	12.37	**216.40**	0.87	**4.06**	39.26	583.04
1434	easy	**91.54**	**100.25**	**65.40**	53.28	72.16	59.02	34.31	1.43	1.85	61.65	226.06
2343	easy	**77.03**	**28.33**	**59.43**	11.99	28.94	28.24	**120.80**	0.84	**4.70**	130.57	816.94
1121	hard	**42.07**	**24.47**	**8.08**	10.93	26.92	4.15	41.53	0.45	0.03	52.32	187.38
1222	hard	**54.06**	**18.32**	**55.89**	9.60	21.04	1.75	26.30	2.19	0.00	62.50	231.83
1242	hard	**4.80**	**58.58**	**21.06**	24.13	17.90	8.75	51.31	1.71	1.82	10.29	388.97
1424	hard	**50.93**	**27.12**	**22.40**	4.54	22.43	5.11	33.89	0.86	0.90	57.61	187.99
1444	hard	**11.75**	**28.30**	**109.18**	20.27	5.65	47.24	73.00	0.05	0.97	52.10	434.94
1141		16.01	36.69	30.87	25.96	8.26	13.80	21.99	0.89	0.80	44.32	452.45
1323		69.11	43.88	111.96	9.81	39.29	15.87	95.60	1.71	0.64	91.31	233.90
2131		51.09	68.68	88.22	28.41	39.23	25.92	17.64	0.56	2.26	16.55	306.97
2141		58.08	12.23	22.79	15.03	18.66	17.95	19.54	0.45	0.83	96.63	639.82
2232		44.05	82.68	188.14	24.22	36.41	12.99	35.71	0.49	1.41	20.66	342.57
2242		49.26	40.26	34.83	33.73	3.14	10.51	25.01	0.47	1.82	72.79	620.80
2434		40.61	73.12	87.80	57.82	49.72	53.92	0.42	0.57	2.75	4.04	414.04
2444		62.69	1.18	86.78	24.81	16.79	42.14	39.11	0.81	1.87	109.71	622.92
3141		109.17	56.45	65.43	13.38	57.89	7.97	1.90	1.01	1.43	113.18	332.85
3242		93.31	42.42	153.31	9.52	39.55	23.49	10.70	0.02	0.41	93.45	278.23
3444		103.30	71.95	174.58	33.01	66.51	11.78	38.69	1.38	0.88	113.75	208.88

Chapter 8. References

ANSI (American National Standards Institute). (1996). *Methods for manual pure-tone threshold audiometry.* New York: American National Standards Institute.

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,* 97-103.

Berkowitz, E., & Berkowitz, S. (2009). Trial Making m files. Personal MATLAB programs.

Best, C. 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). Timonium, MD: York Press, Inc.

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International 5:9/10,* 341-345.

Broselow, E., Hurtig, R.R., & Ringen, C. (1987). The perception of second language prosody. In G. Ioup & S,H. Weinberger (Eds.), *Interlanguage Phonology: The acquisition of second language sound system.* (pp. 350-361). Cambridge, MA: Newbury House Publishers.

Chandrasekaran, B., Krishnan, A., & Gandour, J.T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research, 1128,* 148-156.

Chandrasekaran, B., Gandour, J.T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative neurology and Neuroscience, 25,* 195-210.

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences* (Rev. Ed.). New York: Academic Press.

Duanmu, S. (1999). Stress and the development of disyllabic vocabulary in Chinese. *Diachronica XVI,* 1.1-35.

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, Inc.

Fu, Q.-J. & Zeng, F.-G. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing, 5,* 45-57.

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,* 207-222.

Gottfried, T.L. & Ouyang, G.Y.H. (2006). Training musicians and nonmusicians to discriminate Mandarin tones. *Journal of the Acoustical Society of America, 120,* 3167.

Gottfried, T.L. & Suiter, T.L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of phonetics, 25,* 207-231.

Gottfried, T.L. & Xu, Y. (2008). Effect of musical experience on Mandarin tone and vowel discrimination and imitation. *Journal of the Acoustical Society of America, 123,* 3887-3888.

Grier, J.B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75,* 424-429.

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

Hillenbrand, J.M., and Gayvert, R.T. (2003). Open source software for experiment design and control, *Journal of Speech, Language, and Hearing Research, 48*, 45-60.

Hsieh, L, Gandour, J. Wong, D. & Hutchins, G.D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain and Language, 76,* 227-252.

Jenkins, J.J. (2002). *Signal Detection Theory: A Simple Laboratory Guide*. Unpublished manuscript. Available from author: j3cube@aol.com.

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.

Klein, D., Zatorre, R.J., Milner, B. & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage, 13,* 646-653.

Krishnan, A. (2007). Frequency Following Response. In RF Burkard, JJ Eggermont, M Don (Eds.), *Auditory evoked potentials: Basic principles and clinical application* (pp.313-333). Baltimore, MD: Lippincott, Williams & Wilkins.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research, 25,* 161-168.

Liu, H.-M., Tsao, F.-M. & Kuhl, P.K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology, 43,* 912-917.

Luo X., & Fu, Q-J. (2004). Enhancing Chinese tone recognition by manipulating amplitude envelope: Implications for cochlear implants. *Journal of the Acoustical Society of America, (116),* 3659-3667.

Pallant, J. (2001). *SPSS Survival Manual.* Philadelphia, PA: Open University Press, McGraw-Hill.

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

Shen, X.S., Lin, M. & Yan, J. (1993). F0 turning point as an F0 cue to tonal contrast: A case study of Mandarin tones 2 and 3. *Journal of the Acoustical Society of America, 93,* 2241-2243.

Siegel, S. & Castellan, N.J. Jr. (1988). *Nonparametric statistics for the behavioral sciences.* (2nd ed.). McGraw-Hill: New York.

Tsao, F.-M. (2008). The effect of acoustic similarity on lexical-tone perception of one-year-old Mandarin-learning infants. *Chinese journal of psychology, 50 (2),* 111-124.

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

Wang, Y., Jongman, A., and Sereno, J.A. (2006). Second language acquisition and processing of Mandarin tone. In Li, P., Tan, L.H., Bates, E., and Tzeng, O.J.L.

(Eds.), *Handbook of East Asian Psycholinguistics* (Vol. 1: Chinese). (pp.250-257). Cambridge, UK: Cambridge University Press.

Werker, J.F. & Logan, J.S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37,* 35-44.

Whalen, D.H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49,* 25-47.

Wong, P. (2008). *Development of lexical tone production in disyllabic words by 2- to 6-year-old Mandarin-speaking children.* Unpublished doctoral dissertation, The Graduate Center of the City University of New York.

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,* 420-422.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America,* 95, 2240-2253.

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

Xu, Y. (2001). Sources of tonal variations in connected speech. *Journal of Chinese Linguistics*, monograph series #17. 1-31

Xu, Y. (2007). A Praat script approach to prosody analysis. Downloaded 2007 from: http://www.phon.ucl.ac.uk/home/yi/tools.html

Xu, Y., Krishnan, A. & Gandour, J.T. (2006). Specificity of experience-dependent pitch representation in the brainstem. *NeuroReport, 17,* 1601-1605.

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

Yu, Y. (in prep). *ERP indices of Mandarin tone Processing in monolingual English and bilingual Mandarin-English speakers*. Dissertation in prep., City University of New York Graduate Center.

[i] 4 possible tones for the initial syllable x 4 possible tones for the final syllable =16; but tone 33 always changes to tone 23 due to tone sandhi rules

[ii] All combinations of /dima/ were also recorded, but were not used in this study.

[iii] The trial table for the experiment was created using Matlab routines created for this purpose (Berkowitz & Berkowitz, 2009).

[iv] These blocks of 12 are actually 6 unique trials repeated twice each; that is, when you reverse the order of delivery for the same pairs, the result is an identical list.

[v] 15 disyllables * 14 disyllables = 210. Because order is irrelevant, 210/2 = 105.