CROSSLINGUISTIC PERCEPTION OF PITCH IN LANGUAGE AND MUSIC

by

Evan David Bradley

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Linguistics and Cognitive Science

Fall 2012

CROSSLINGUISTIC PERCEPTION OF PITCH IN LANGUAGE AND MUSIC

by

Evan David Bradley

Approved: ______________________________

Benjamin Bruening, Ph.D.
Chair of the Department of Linguistics and Cognitive Science

Approved: ______________________________

George H. Watson, Ph.D.
Dean of the College of Arts and Sciences

Approved: ______________________________

Charles G. Riordan, Ph.D.
Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Jeffrey Heinz, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Irene Vogel, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

H. Timothy Bunnell, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

L. Robert Slevc, Ph.D.
Member of dissertation committee

The quality of our perceptions determines the quality of our judgment.
Our judgment determines how we interact with the world.
How we interact with the world changes the world.
Therefore, the quality of our perceptions changes the world we perceive.

(Robert Fripp)

ACKNOWLEDGEMENTS

This project would not have been completed without the guidance of assistance of many people. I would like to thank them here, in no particular order.

My advisors, Jeffrey Heinz and Irene Vogel, were invaluable to completion of this dissertation from beginning to end. Irene was crucial to helping me crystallize my ideas around the topic, and Jeff's combination of scientific and practical advice, along with gentle encouragement made all the difference. The other members of the committee were extremely helpful as well. Timothy Bunnell provided advice during the development of the project, and helpful guidance during the data collection and analysis. I thank Robert Slevc for his insightful comments and perspective, as well as his flexibility.

The members of the Phonology and Phonetics Laboratory were very helpful in developing the project from the beginning, especially continuous support from Regine Lai, Tim O'Neill, and Jane Chandlee. Others who provided helpful commentary or technical guidance along the way include Jean-Philippe Laurenceau, Arild Hestvik, Suzanne Burton, Terry Gottfried, William Idsardi, Anirrudh Patel, Richard Ashley, Jennifer Alexander, Patrick Wong, and members of the Northeast Music Cognition Group.

Financial support for the project came from the University of Delaware and the National Science Foundation (DDRIG BCS-1156289). I thank NSF, the UD Office of Graduate and Professional Education, and the UD Research Office for their support, as well as Joan Maling, William Badeker, and several anonymous reviewers.

The stimuli used in the studies would not have been constructed without the assistance of Zenghong Rebecca Jia, Jiao Wang, Yaping Tsai, Yu-Lin Chiu, Chiao-Yu Tuan, and Tomilayo Jacobs. Several University of Delaware instructors were kind enough to allow me into their classrooms to collect data: Maria Tu, Daniel Stevens, Lauren Ketter, Zhiyin Renee Dong, and Iris Busch. The Phonology and Phonetics Lab research assistants, Thomas Julian and Megan Rosales, were instrumental in running the experiments.

Recruiting the diverse group of participants who took party in the study would not have been possible without the help of many people, especially Jaime Resnick, Kenneth Hyde, Romuladus Azuine, Elizabeth Adetobi-Oladele, Sussan Ekejiuba, Akinbiyi Akinlabi, Emmanuel Iwuoha, Adeboye Adejare, Adebanke Adeyemo, Paul Head, Robert Schrader, The Xi Mu chapter of Phi Mu Alpha Sinfonia, and The University of Delaware Fightin' Blue Hen Marching Band.

I am sure to have left some names off of the list, and I sincerely apologize for that. as well as for any remaining errata.

for Laura, who showed me how to do it;
for Liam, who convinced me I could do it;
and for Quinn, who made me finish it.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AP Absolute Pitch

bpm beats per minute

ERP event-related potential

FFR frequency following response

fMRI functional magnetic resonance imaging

IRN iterated rippled noise

ISI inter-stimulus interval

L2 second language

MDS multidimensional scaling

MEG magnetoencehalography

MET Musical Ear Test

MMN mismatch negativity

PSOLA Pitch-Synchronous Overlap and Add

RHT Reverse Hierarchy Theory

S/N signal-to-noise

SSCLMH Shared Sound Category Learning Mechanism Hypothesis

ABSTRACT

This dissertation investigates the ways in which experience with lexical tone influences the perception of musical melody, and how musical training influences the perception of lexical tone. The central theoretical basis for the study is a model of perceptual learning, Reverse Hierarchy Theory (Ahissar et al., 2009), in which cognitive processes like language tune neural resources to provide the sensory information necessary for the perceptual task; these sensory resources are then available to other cognitive processes, like music, which rely on the same perceptual properties. This study proposes that the tone properties *pitch height*, *pitch direction*, and *pitch slope* correspond to the melodic properties *key*, *contour*, and *interval*, respectively, and this correspondance underlies crossover effects between lexical tone and melody perception.

Specifically, the study asks three questions:

1. whether differences in melody perception between tone and non-tone language speakers, and among speakers of different tone languages, can be linked to specific properties of the languages' tonal inventories;

2. whether melody perception is affected by second language experience with a tone language; and

3. whether musical ear-training leads to enhanced perception of lexical tone.

To address (1), a standardized test of music perception (the Musical Ear Test; Wallentin et al. (2010)) was administered to tone (Mandarin and Yoruba)

and nontone (English) language speakers. Tone language speakers demonstrate more accurate melody perception than English speakers; rather than a uniform advantage, however, this effect is limited to those specific properties argued to be shared between language and music. Further, Mandarin and Yoruba speakers do not perform identically on melodic perception, suggesting linguistic effects on melody perception are related to differences between the tonal inventories of the languages.

Attempts to extend this hypothesis to second-language tone experience (2) were not successful; Mandarin learners did not perceive melody similarly to native speakers. Further study with more proficient second language speakers is necessary.

The role of explicit perceptual music training (3) was examined by assessing the effects of aural skills training on musicians' perception of Mandarin lexical tones. The results reveal that this training did not lead to improvement in the perception of these tones in a similar fashion to native or second language speakers of Mandarin, but did change musicians' response bias toward the tones in a manner consistent the general hypothesis.

This work attempts to better understand pitch perception within a theoretical framework of perceptual learning. Taken together, the results partially support the specific proposed mappings between structural properties of language and music, and more generally support a framework for explaining these and other cases of crossover between language and music. These findings address questions of cognitive modularity and the relationship between language and music, as well the role of sensory experience during development and adulthood.

Chapter 1

INTRODUCTION

1.1 Purpose

Language and music each draw on rich acoustic systems, and represent two of the few types of cognition thought to be truly unique to humans. Much has been made of their similarities, but the degree to which they are connected in the mind is far from fully understood.

This research focuses on *pitch* as a key element of both language, in the form of lexical tones, and of music, in the form of melody. Perceptual links between the two systems are investigated; specifically:

1. how general auditory mechanisms for pitch perception are used by musical and linguistic systems;

2. how such perceptual mechanisms change as a result of linguistic and musical experience; and

3. whether and how these mechanisms and their responses to experience result in a transfer of knowledge about pitch components between language and music.

The influence of tone language experience on musical melody perception is examined in relation to

1. the tonal features of the language, and

2. whether the language is acquired natively or learned in adulthood.

These are considered through examination of native speakers of Mandarin and Yoruba, along with adult learners of Mandarin.

The influence of musical experience on tone language perception is examined through consideration of English-speaking musicians undergoing aural skills training, and the effects this training has on the perception of Mandarin lexical tones.

1.2 Hypotheses

The specific hypotheses to be tested in each experiment stem from a General Hypothesis based on theories of perceptual learning. This framework states that learning to perceive a complex acoustic domain, such as language or music, results in perceptual tuning of the auditory system to specific acoustic properties relevant to the task.

Generalization of experience across domains occurs not due to extension of linguistic or musical knowledge, or a general enhancement of auditory processing, but due to changes in the encoding of dynamic pitch, which corresponds to structural elements of language and music.

Specifically, it is anticipated that tone and non-tone language speakers will differ on musical tasks as a function of the tonal features of their language; likewise, musicians will differ in tone perception from those without musical training based on their greater sensitivity to elements of musical structure.

1.3 Outline

Chapter 2 begins with a review of previous research on tone perception in several languages, including typological and phonetic descriptions of tone systems. Crosslinguistic studies demonstrating effects of language experience on the perception of tones and non-speech sounds illustrate the levels of representation relevant to the encoding of linguistic pitch.

Chapter 2 continues with a review research on the perception of pitch in music, with an emphasis on drawing parallels between the characteristics of musical and linguistic pitch systems, including the effects of native musical culture and training on perception, and the perception of and memory for melody in relation to more basic units of musical pitch.

Evidence is presented to establish perceptual links between music and language, including data from behavioral, neurophysiological, and second language learning studies, which are discussed in relation to theories of perceptual learning.

Gaps in the existing literature are identified, including:

1. unresolved questions about the effect of tone language experience on the perception of musical pitch, specifically the lack of explicit links between structural properties of language and music giving rise to these effects;
2. insufficient distinction between the effects of different kinds of musical experience, specifically between the participation in or performance of music and perceptual (ear) training; and
3. a shortage of crosslinguistic tone/melody perception studies including speakers of *register* tone languages.

A General Hypothesis is described which lays the groundwork for a series of more specific hypotheses about the consequences of linguistic and musical experience for domain-general pitch perception, culminating in a mapping between perceptual properties of lexical tone and melody (Table **??**). This mapping is formulated by integrating the available evidence with the perceptual learning models discussed earlier. A series of experiments to test these hypotheses is proposed.

Chapter 3 describes an experiment comparing melody perception by native speakers of three languages: English, Mandarin, and Yoruba. These languages represent a diverse sample of the tonal properties found across languages—English is a nontonal language; Mandarin has lexical contour tones, and Yoruba is a register tone language, with level lexical tones. The languages also generate different predictions based on the mapping between linguistic and melodic pitch properties (Table **??**) developed in Chapter 2.

The results of Experiment 1 support the General Hypothesis, but only partially support the specific hypotheses about English, Mandarin, and Yoruba based on the tone–melody mapping. New questions arising from these results and potential revisions to the mapping are discussed.

Chapter 4 describes two attempts to extend the results of Experiment 1 from native to second language experience. The melodic perception ability of adult second language learners of Mandarin is examined over time (Experiment 2) and in cross-section (Experiment 3).

The results of these experiments fail to support the extension of the General Hypothesis to second language tone experience. This null result is likely due to limitations in the methodology and samples employed in these studies—the specific hypotheses about tone language learning should not be abandoned without further study.

Chapter 5 considers whether musical aural skills (ear) training can influence the perception of lexical tone in the absense of linguistic training. Experiment 4 compares Mandarin lexical tone perception by musicians before and after aural skills training.

Interestingly, the results indicate that aural skills training changes the way Mandarin tones are perceived, but not in a native-like way. The nature of these changes are argued to support the General Hypothesis and the level of representation at which the tone–melody mapping is formulated.

Chapter 6 discusses the results of experiments presented in Chapters 3–5 in relation to one another and the General Hypothesis and specific links between tone and melody developed in Chapter 2. Implications of the findings for theories of perception, language learning, and music education are discussed, and future avenues of research are identified.

Chapter 2

BACKGROUND

This chapter gives an overview of research on pitch perception in language and music, focusing on parallels and overlaps between the two (Section 2.1). Cases of crossover effects between tone language and music are discussed in detail, and a framework for general hypotheses will be developed based on perceptual learning theories (Section 2.2). Finally, outstanding questions are identified and a series of experiments to address them are formulated (Section 2.3), to be described in Chapters 3–5.

2.1 Summary of Previous Work

Confusion can result from the use of the terms 'pitch' and 'tone' in different contexts, and a distinction must be made between physical, perceptual, and systematic descriptions of auditory phenomena. Yip (2002) succinctly describes the levels that must be considered in regard to lexical tone: *Frequency* (F_0) is a physical property of the acoustic signal, *pitch* is the perception of frequency and other acoustic properties by a listener, and *tone* is an abstract linguistic object. McDermott and Oxenham (2008) described pitch as "the perceptual correlate of periodicity in sound", and this perceptual object is available to multiple cognitive systems, including language (*e.g.*, tone, intonation) and music (*e.g.*, melody, harmony). The term 'note' will be used to denote the musical equivalent of linguistic tone; that is, a musical category based on pitch.

2.1.1 Tone

Pitch is used in several ways in language. This review will focus lexical tone, which is the use of suprasegmental pitch information to distinguish individual words. This is different from intonation, which is the use of pitch to distinguish sentence types. An intermediate case is the limited use of pitch to distinguish words in a manner similar to stress, or pitch accent (Maddieson, 2005). Thus, the degree of lexical pitch-use in a language may be considered as a continuum, from intonation-only languages at the less-tonal end, to full-fledged tone languages (and from those with few to those with many tones) at the more-tonal end.

2.1.1.1 Typology

By some estimations, as many as 60–70% of the world's languages can be classified as tonal (Yip, 2002), but only a portion of these have large tonal inventories (Maddieson, 2005). Tonal languages are clustered geographically, with the majority occurring in East and Southeast Asia, Africa, and America (Maddieson, 2005). Most tone languages (up to 80%) contain only relatively level tones (Maddieson, 1978) (known as register tone languages, common in Africa), and phonemic contour (rising, falling, concave, or convex) tones tend to appear only in languages with a relatively large number of tones (Maddieson, 2005) (contour tone languages, common in East and Southeast Asia). An example of a register tone language with three tones (Low, Mid, and High) is Yoruba; an example of a contour tone language is Mandarin, which has four tones, including a high-level, falling, rising, and dipping tone.

Many register tone languages have a simple binary contrast between low and high level tones, but register tone languages may contrast up to five levels, with three being common, and increasing rarity for each additional tone (Yip, 2002; Gussenhoven, 2004). Phonetically, tones in register tone languages need not be perfectly flat, but must be at least flat enough that, as Maddieson (1978) described, "a level pitch is an acceptable variant". Rising or falling tones may occur in register tone languages as a result of phonological processes, but they do not have the same phonemic status of rising or falling tones in contour tone languages (Yip, 2002).

Register tones are defined by their position within the pitch space used by the speakers. Rather than maximal dispersion of tones within the pitch range of a speaker, such that a language with two tones will have a greater distance between its tones than a language with three tones (with each language having a similar distance between its lowest and highest tones), in most register tone languages, tones are separated by 2–3 semitones, with the total pitch range increasing for languages with more tones (Maddieson, 1978, 1991).

By some accounts, contour tone languages may have up to eight or even thirteen tones (Yip, 2002; Patel, 2003), though segmental, syllabic, and voice quality correlates make determining the exact number of phonemic categories difficult (Gussenhoven, 2004). As with register tone languages, contour tone languages with a greater number of tone contrasts are rarer than those with fewer, and complex contours (concave or convex) are rarer than simple rising and falling tones (Yip, 2002). Despite their name, contour tone languages also contain level tones; in fact, if a language contains a phonemic contour tone, this implies that it has at least one level tone (Patel (2008b); Maddieson (1978) notes a few possible exceptions). Contour tones are rare in languages with only three tones, suggesting that contour tones result when an upper bound on level tones is reached (Yip, 2002; Patel, 2008b).

2.1.1.2 Perception, Categorization, and Normalization

Like the perception of other linguistic objects, the perception of tone is a complex process which is influenced by characteristics of the speech signal, speaker, and listener.

Categorical perception, or the perception of continuous physical dimensions in a discrete fashion, is found throughout language, and its hallmark is the invariance of perception to irrelevant differences between tokens (Liberman, Harris, Hoffman,

& Griffith, 1957). Like segmental phonemes, tone is a linguistic category which must accomodate variation from token to token. Even within an utterance, tones of the same type may not have the same F_0 or shape. Variance occurs for several reasons in different languages, including :

- coarticulation with neighboring tones, resulting in *peak delay*, or the shifting of F_0 targets within the syllable or onto the following syllable (Y. Xu, 1999a, 1999b, 2001; Y. Xu & Wang, 2001); this process has been phonologized in Yoruba, resulting in rising and falling tones derived from an underlying level-only tonal inventory (Akinlabi & Liberman, 2001). Indeed, acquired knowledge of the acoustic effects of coarticulation may aid in tone identification in context (Gottfried & Suiter, 1997; Y. Xu, 1994).

- *downtrend*, which as used by Connell and Ladd (1990), refers generally to any lowering of pitch across an utterance, including *declination*, is a lowering of F_0 from beginning to end of an utterance resulting from lowered air pressure in the lungs, and associated phonological processes such as *downdrift* and *downstep*, which result in the lowering of tones in certain environments. This can cause overlap of tonal categories, because a high tone at the end of an utterance may in fact be lower than a low tone at the beginning of the utterance. For example, in Yoruba, the deletion of a low tone can cause downstep on a following high tone (Connell & Ladd, 1990). Other languages, termed *discrete-level* languages, are less subject to various downtrend processes, keeping their tone categories separate (Connell & Ladd, 1990; Patel, 2008b).

- intonation, requiring listeners to integrate prosodic context when identifying tones (Connell, Hogan, & Rozsypal, 1983; S.-h. Peng, 1997).

While some have claimed that tone perception is not categorical (Abramson, 1975, 1979) in all of the ways originally described by Liberman et al. (1957), more recent studies suggest that task and stimulus design influence the nature of tone perception seen in experiments, and that tone perception is categorical in many tasks; for example, Francis, Ciocca, and Ng (2003) compared the perception of Cantonese tones in an identification task to a discrimination task, finding that perception of the tones is more categorical during identification than during discrimination. Some studies have suggested that intonational pitch contours may be perceived categorically (Köhler, 1987; Pierrehumbert & Steele, 1989), which may indicate that even speakers of non-tonal language like English have experience processing linguistic pitch categorically

The acoustic properties of speech categories vary not only within but between talkers. Listeners must compensate for variation within talkers, and for differences between talkers, a process known as *speaker/talker normalization* (Johnson, 2005). Normalization occurs for a wide range of acoustic properties of speech, but it is particularly relevant to tone perception because the pitch range used by speakers can vary by a great deal based on gender, anatomy, affect, and other factors, such that a "high" tone spoken by male speaker may have a lower F_0 than a "low" tone spoken by a female speaker (C.-Y. Lee, 2009).

If a tone has a distinctive contour, it may be identified in isolation, but in the case of level tones, or of tones with similar contours at different ranges, information about the pitch range used by the talker is necessary to identify the tone (Wong & Diehl, 2003). This information can be deduced from external cues, such as the preceeding utterance context (Y. Xu, 1994, 1997; Moore & Jongman, 1997; Wong & Diehl, 2003), or internal cues: C.-Y. Lee (2009) demonstrated that Mandarin listeners were able to identify isolated tones from multiple talkers, suggesting that they could normalize the speakers' pitch ranges based on internal characteristics of the tone stimuli. Honorof and Whalen (2005) demonstrated that even English speakers can place the pitch of an isolated vowel within the range of unknown speakers, suggesting that they are sensitive to acoustic correlates of pitch other than F_0.

2.1.1.3 Multidimensionality

Tone is not a unitary phenomenon, but is an abstract linguistic object composed of multiple perceptual properties, including not only F_0, but spectral (*e.g.*, harmonics, voice quality) and temporal (*e.g.*, duration, amplitude, rise/fall time) components; listeners employ these properties as perceptual cues to tonal categories to different degrees based on the relative importance of the dimension in their native tone system (C.-Y. Lee, 2009). These multiple cues provide a degree of redundancy, allowing accurate perception in noise (Kong & Zeng, 2006) or of degraded stimuli, such as in whispered speech where F_0 information is unavailable, or speech which has been high-pass filtered to remove F_0 information (Liang, 1963; Fu, Zeng, Shannon, & Soli, 1998; Abramson, 1973). Liu and Samuel (2004) found that Mandarin speakers emphasize such secondary tonal cues when they know that primary cues will be unavailable, for example, when whispering.

The relative importance of various perceptual cues does not only vary between languages, but between particular tones or contexts within languages (Connell, 2000). For example, Fu et al. (1998) found that Mandarin tones 3 and 4 could be reliably recognized using only temporal cues with spectral information removed, while recognition of tones 1 and 2 suffered from removal of spectral cues.

Although pitch, and by extension, F_0, is the primary component of tone, the F_0 content of tones is also perceived along several perceptual dimensions. This multidimensionality provides a further degree of redundancy to the tone system (Gottfried & Suiter, 1997; C.-Y. Lee, 2009), but languages differ in the weight afforded to the various perceptual properties of F_0.

Gandour and Harshman (1978) used multidimensional scaling (MDS) to identify factors involved in tone perception by speakers of different languages. Thai, Yoruba, and English speakers were asked to rate the similarity of synthesized words which contained a variety of level, rising, and falling pitch patterns. Their analysis revealed several key subcomponents of pitch perception, which resemble phonological features of tone described in acoustic terms:

- Average pitch (henceforth *height*), which distinguishes tones based on their average F_0 level; thus, it maximally distinguishes high (55) from low (11) tones, and groups tones like 15 and 51 together with 33, because the average pitch of the rising, falling, and level tones is the same.
- *Direction*, which distinguishes rising, falling, and level tones, regardless of their pitch range or degree of pitch change; thus, it treats a low-rising tone (13), a high-rising tone (35), and a low-to-high rising tone (15) as similar, and distinct from level and falling tones.
- *Slope*, which distinguishes tones based on the steepness of pitch change; thus, it groups tones which change pitch rapidly (15, 51), distinguishing them from tones which do not change at all (11, 33, 55), with less steeply changing tones (35, 53) in between.
- *Length*, which distinguishes syllables based on duration, a common correlate of tone in many languages.
- *Extreme endpoint*, which distinguishes tones which end in the extremes of the pitch range (1 or 5) from tones which end in the middle (3); thus, it groups tones like 11, 15, and 55 together, in opposition to tones like 33 and 53.

These properties were classified as either static properties (*e.g.*, *height*, *endpoint*, *length*) or dynamic properties (*e.g.*, *direction*, *slope*). Gandour and Harshman (1978) argue that static properties reflect general auditory capabilities, while dynamic properties reflect language-specific dimensions (they note that *height* could reflect either language-specific or general auditory capabilities, or a combination of both).

Height was found to be the most important property for determining the similarity of tones for all three language groups, but was most important to English speakers, while *direction* was less important to English speakers, relative to the other language groups. The authors suggest that the English group, as speakers of a non-tonal language, were reliant primarily on non-linguistic static properties.

Thai and Yoruba speakers were again distinguished from English speakers by the *direction* and *slope* properties; these dynamic properties were more important for the tone language speakers than for the English speakers. The two tone language groups did not differ on the weight given to *direction*.

The authors attribute the importance of *direction* in Yoruba to the presence of rising and falling "allotones" in Yoruba; that is, Yoruba contains tones with surface (but not underlying) contours, therefore the detection of rising or falling pitch is important for Yoruba speakers. Thai and Yoruba speakers differed on the importance given to *slope*, with Thai speakers using this property moreso than the Yoruba speakers, which the authors speculate may be due to presence of underlying contour tone categories in Thai, while the surface contours in Yoruba arise only due to phonological rules. It is also consistent with descriptions of the Yoruba tonal inventory; although Yoruba has rising and falling tones, it does not have multiple tones with either rising or falling elements, thus de-emphasizing the importance of *slope* in distinguishing individual tones, though it may still play a role in tone perception in continuous speech (Yu, 2009). Yoruba speakers in turn displayed more sensitivity to *slope* than did English speakers, indicating the importance of these perceptual properties is gradient across language

In summary, the analysis by Gandour and Harshman (1978) revealed that not only does language background influence the salience of various properties of pitch, but the relative importance of these properties can be linked to characteristics of the tonal system of the speakers. Their results suggest that register and contour tone languages share a reliance on dynamic properties of F_0 which non-tone languages do not, but differ in the particular importance assigned to individual dynamic properties of pitch. Although it was suggested that English speakers rely only or mostly on general auditory abilities, rather than language-specific properties when perceiving tone, the possibility that other kinds of prosodic categories influence tone perception by non-tone langauge speakers cannot be entirely ruled out (Francis, Ciocca, Ma, & Fenn, 2008; Wang, Spence, Jongman, & Sereno, 1999).

Gandour (1983) further investigated the influence of language experience on the perception of linguistic pitch by comparing judgments of synthesized continuua of tones by speakers of four Asian contour tone languages (Mandarin, Cantonese, Taiwanese, and Thai) to English. Each of these languages contains a different number and configuration of tones. From similarity judgments of a set of nineteen synthesized tones, the two most significant dimensions were extracted and impressionistically labelled by the authors as *height* (a static property) and *direction* (a dynamic property) Gandour (1983) is careful to note that in MDS studies, the particulars of the stimulus set and task exert an influence on the nature and quantity of dimensions found; however, across studies (Gandour & Harshman, 1978; Gandour, 1978, 1979a, 1979b, 1983; Avelino, 2003; Khouw & Ciocca, 2007; Chandrasekaran, Gandour, & Krishnan, 2007) general correspondances between important dimensions are found, suggesting that these dimensions are psychologically real, rather than experimental artifacts. Thus, Gandour (1983) interprets his dimensions, *height* and *direction* to correspond to *average pitch* and *direction* as described in Gandour and Harshman (1978), respectively, and verified that the distinctions made by *height* correspond to the average frequency values of the tones.

Similar to the findings of Gandour and Harshman (1978), *height* was the most important property across all language groups. The Thai group relied less on *height* than any the other groups, and English and Cantonese relied more heavily on *height* than the other groups. In fact, the English and Cantonese groups were indistinguishable in the importance placed on height.

English and Cantonese speakers did differ in their use of *direction*. English speakers relied on *direction* the least of all the groups, while Thai speakers placed the greatest load on direction (in fact, for some it was more important than height). English speakers relied on *direction* the least of all the groups, while Thai speakers placed the greatest load on *direction*; the Chinese language groups fell in the middle and were not different from one another with respect to direction. Gandour (1983) suggests that the difference between the Thai and Chinese groups on *direction* arises from differences in the phonology of the two language families (specifically, the fact that the Chinese languages contain tone sandhi rules, while Thai does not) rather than differences in tonal inventory, also noting that in Gandour and Harshman (1978), Thai and Yoruba patterned together on most dimensions, including *height/average pitch* and *direction*; these two observations, taken together, form the basis for his assertion that the importance of particular properties of pitch to perception does not seem to correlate with the typological status of a language (*i.e.*, register or contour), but rather with its phylogenetic origin (*i.e.*, Chinese vs. Thai) (*cf.* the speculation by Gandour and Harshman (1978) that differences between Thai and Yoruba on the *slope* dimension could be related to the register/contour distinction).

The dimensions and the effects thereon by native language revealed by these studies are consistent with other findings on tone perception in the languages investigated by Gandour (1983). For example, (Chandrasekaran, Krishnan, & Gandour, 2007b) found that the event-related potential (ERP) responses of Mandarin and English listeners can be predicted by differences in the weighting of these cues.

Even though speakers of different languages rely on pitch dimensions to different degrees, these cues can still be used redundantly. For example, although pitch movement is a primary cue to tone in Mandarin, Mandarin speakers can identify tones based on truncated stimuli which include only six glottal pulses, eliminating dynamic F_0 information, suggesting that static properties like onset F_0 height (and possibly other cues, like voice quality) can also be used by Mandarin speakers to identify contour tones (Gottfried & Suiter, 1997; K. M. Lee, Skoe, Kraus, & Ashley, 2009).

Conversely, Abramson (1978) notes that the so-called "level" tones of Thai contain some pitch movement, and used synthesized continua of Thai tones to demonstrate that this pitch movement aids in tone identification beyond that afforded by completely static pitch height, although truly static tones can still be identified by Thai speakers (Maddieson, 1978). (House, 1989) found that the relative ability to perceive pitch level and pitch movement is affected by syllabic context, suggesting that the redundancy of cues might aid the perception of tone in different syllable types.

Research on the relevant perceptual cues to tone for non-Asian languages beyond the study including Yoruba by Gandour and Harshman (1978) is more scarce. Connell (2000) examined the perception of tone in Mambila, a Benue-Congo language with four level tones. Phonetically, Mambila tones are very level, although the highest and lowest tones are reported to have a slight rise and fall, respectively. The results indicated that some Mambila speakers can determine the tonal identity of monosyllables using pitch height alone, although their assignment of particular frequencies to words was not uniform across all tones; that is, some tone categories included a greater range of F_0 than others, suggesting other cues may be involved for some contrasts.

Connell (2000) compared these results to those of English speakers, who classified the stimuli in a categorical fashion, but partitioned the pitch space in a more uniform way than Mambila speakers; that is, each category included a similar size range of F_0. Importantly, a signal detection analysis revealed no difference in sensitivity to pitch between the Mambila- and English-speaking groups. The results of these experiments suggest that although the Mambila- and English-speaking groups could perceive the continuous dimension of pitch equally well, the Mambila speakers used their language-specific knowledge to map F_0 onto their tonal categories, while the English speakers' perception of pitch was not influenced by this knowledge (requiring them to rely on other kinds of knowledge, or to create ad-hoc, stimulus-driven categories).

Avelino (2003) examined the perception of tone continua in Yalálag Zapotec, an Otomanguean language of Mexico which has High, Low and Falling tones, finding that Zapotec listeners are sensitive to average F_0 height, F_0 endpoints, and F_0 direction, but employ these cues differently for each tone contrast.

In summary, tone perception appears to be implemented crosslinguistically by several perceptual properties of F_0, which can be classified as *static* or *dynamic*. An important static property is *height*, and two important dynamic properties are *direction* and *slope* of pitch change. The way in which these dimensions are used is determined in part by general principles of audition, the inventory of tones in the language of the perceiver, and possibly the phonology of the language. Tone language speakers can be distinguished from speakers of non-tone languages by their greater sensitivity to dynamic pitch cues. Non-tone language speakers, by contrast, rely heavily on static pitch cues, such as average pitch height. This characterization is supported by findings of more accurate neural encoding of dynamic pitch information by speakers of tone languages compared to speakers of non-tone languages (Krishnan, Xu, Gandour, & Cariani, 2005; Chandrasekaran, Krishnan, & Gandour, 2007b, 2007a; Chandrasekaran, Gandour, & Krishnan, 2007; Krishnan, Gandour, &

Bidelman, 2010b).

Regarding the register/contour distinction, Gandour and Harshman (1978) suggest that what may distinguish contour tone from register tone language speakers is that although both language types contain surface contours, the underlying flat tones of register languages cause speakers to attend primarily only the *direction* of pitch changes, while the underlying contours of contour tone languages cause those speakers to attend to the *slope*, or rate of change, of pitch in addition to its direction. It is important to note that few studies have compared register and contour tone languages directly, so this typological generalization is speculative at best.

2.1.1.4 Native and Second Language Tone Perception

Evidence of this acquired sensitivity to the perceptual and systematic properties of a native tonal system can be seen in the perception and learning of nonnative tones. As with other aspects of language, as listeners acquire the distinctions relevant to their native language, their perceptual systems adapt in order to store and process native input, and the perception of nonnative linguistic elements changes as a function of native categories. The effect of experience-induced sensitivity to perceptual properties of F_0 is manifested not only in the perception of synthesized continua of tones, but in the perception of nonnative tone systems. This results in perceptual difference between tone and non-tone language speakers, and between speakers of different tone languages.

In addition to the studies cited above detailing differences in the weighting of perceptual cues to tones by speakers of different language, some other studies suggest that the brains of tone language speakers process tones in linguistically relevant ways, while those of non-tone language speakers do not. Gandour (1998) found that Thai speakers show activation near Broca's area when processing words based on tone, while English speakers do not, and Mandarin speakers have a right ear/left hemisphere advantage for tone processing, while English speakers do not

(Wang, Jongman, & Sereno, 2001). This is compatible with traditional theories of hemispherization which suppose that lexical and syntactic processing occurs primarily in the left hemisphere, while prosodic processing occurs primarily in the right hemisphere. Under these theories, non-tone language speakers do not process pitch information in a lexically relevant way, and so do not show the same hemispherization as speakers for whom tone is lexically meaningful. Wang, Sereno, Jongman, and Hirsch (2003) used functional magnetic resonance imaging (fMRI) to show that as English speakers learned Mandarin tones, they recruited new areas of cortex, especially in the left hemisphere, becoming more like tone language speakers, who process tone to a greater degree in the left hemisphere, while naïve English speakers process tone along with other pitch tasks in the right hemisphere (Klein, Zatorre, Milner, & Zhao, 2001).

2.1.1.5 Levels of Representation

In order to perceive the sounds, including tones, of a language, phonological categories must be formed (Wang et al., 1999; Hallé, Chang, & Best, 2004; G. Peng et al., 2010). The formation of such categories seems to be necessary before pitch can be used to distinguish words in a foreign language (Wong & Perrachione, 2007; J. Lee, Perrachione, Dees, & Wong, 2007; Chandrasekaran, Sampath, & Wong, 2010). With regard to tone, this requires listeners to accurately perceive and attend to the perceptual properties of F_0 relevant to the categories in question (Wayland & Guion, 2003, 2004; Francis et al., 2008), which Wong and Perrachione (2007) describe as a "phonetic–phonological–lexical continuity," in which lower-level acoustic knowledge must be established before forming phonological categories which can be used to contrast word meanings. Wong and Perrachione (2007) found that English-speaking learners' pre-existing ability to identify the relevant pitch contours predicted their success at identifying Mandarin words based only on tone; performance was also correlated with degree of musical training. Learners with higher pre-training pitch perception ability can also accomodate greater stimulus variability during training J. Lee et al. (2007), and better learners attend more to pitch direction than poorer learners Chandrasekaran et al. (2010). In the terminology employed by Wong and Perrachione (2007), the more successful learners are further along the phonetic–phonological–lexical continuum because they can perceive the relevant acoustic properties better than other learners, and therefore can more quickly create new phonological categories.

The effect of linguistic experience is evident at each level of this continuum, and it appears that perceptual tuning may extend in a top-down fashion to the level of general acoustic perception. Thus, there are are at least three levels of representation for pitch that must be considered:

1. a pre-linguistic *acoustic level*, at which acoustic information is extracted and encoded;

2. a *phonetic level*, at which phonetic features are recognized;

3. a *phonological level*, at which phonetic dimensions are mapped to categorical representations in memory.

The effects of linguistic experience on each level will be discussed below, in reverse order.

2.1.1.5.1 Phonological

A key question about nonnative tone perception is to what degree native tone categories either help or hinder perception and learning of new categories. Theories of segmental acquisition and second language (L2) perception (Best, McRoberts, & Goodell, 2001; Flege & Liu, 2001; Iverson, Kuhl, Akahane-Yamada, & Siebert, 2001), although they differ in their details, generally predict that a listeners' native speech categories can both help and hinder perception of new speech sounds by linking them to familiar ones, either easing learning by reducing new information, or interfering by making spurious connections which distort the categories of the L2 system. Categorical effects also appear to be active in the domain of tone.

It has been suggested that speakers of non-tone languages do not perceive pitch in the same categorical fashion as tone language speakers. Hallé et al. (2004) found that Taiwan Mandarin Chinese speakers perceived synthesized continua of Mandarin tones categorically (*i.e.*, greater performance for between- versus within-category comparisons), while French speakers, who perceived contrasts better than chance, did so similarly across the Mandarin categories. They note that "the acoustic correlates of tones, F_0 and intensity contour, are used in French, just as in any language, at the sentential intonation level. Tone contours thus are not completely irrelevant to a French ear with respect to their putative linguistic value. In that sense, we cannot consider tone contrasts to be [not assimilable] for listeners of French (or any other language)." The cause of French listeners deficit, according to the authors, is not that French listeners cannot perceive tones categorically at all, but that the Mandarin tones are sufficiently different from their native intonational categories as to be too difficult to assimilate. The authors go so far as to say that the French listeners are not only not perceiving the phonemic categories, but that they are perceiving the tones as "nonlinguistic melodic variations".

G. Peng et al. (2010) compared the categorical pitch perception of Mandarin, Cantonese, and German speakers as they heard rising and falling continua of speech and nonspeech stimuli. The non-tone (German) language speakers did show evidence of categorical perception, but their category boundaries were not as sharp as those of the tone-language speakers; in addition, the location of category boundaries among the Chinese groups seemed to be influenced by the particular tones of their native language.

Evidence contrary to the idea that nontone language speakers do not perceive speech pitch categorically comes from Wang et al. (1999), who trained English speakers to perceive Mandarin tones in a lexical identification task using a high-variability training set. The learners could generalize the learned tone categories to new (untrained) stimuli, and retained their training for at least six months, suggesting they had formed and maintained new categories. Although participants improved in identifying all four tones, the contrast between Tone 1 and Tone 4 was reported to be the most difficult, which the authors suggest is due to interference with the English-speaking participants' native intonational categories; while Tones 2 and 3 are acoustically similar, and are easily confusable (Shen & Lin, 1991), they improved greatly after training, while the contrast between Tones 1 and 4 were resistant to improvement, perhaps because they both correspond to English segmental level stress intonation. Wang et al. (2003) demonstrated that this perceptual training subsequently improved these speakers' accuracy in production, as well; speakers improvement was not uniform, however—pitch contour accuracy improved to a greater degree than did pitch height, supporting the distinction between these properties found by earlier studies (Gandour & Harshman, 1978; Gandour, 1983, *inter alia*).

2.1.1.5.2 Phonetic

Another important question is whether speakers of different languages differ in their ability to use phonetic, as well as phonological, knowledge to perceive tone. Wayland and Guion (2003) found, unsurprisingly, that experienced English learners of Thai demonstrate better perception of Thai tones than do English speakers with no experience with Thai. Experienced English learners of Thai also show an effect of inter-stimulus interval (ISI), with improved perception of the most difficult contrast with shorter ISI, a pattern not shown by native Thai controls. Based on a link between shorter ISI and a "phonetic mode" of perception and longer ISI with a "phonological mode" Werker and Tees (1984); Burnham et al. (1996), the authors suggest that this indicates that while the experienced English group demonstrated improved phonetic perception compared to naïve listeners, they still lacked the ability to use phonemic categories from long term memory to encode the stimuli phonologically.

Wayland and Guion (2004) trained English and Chinese (Mandarin and Taiwanese) speakers to perceive Thai tones. Chinese speakers discriminated the Thai tone contrast better than the English group at a short ISI even before training; after training, the Chinese group was better than the English group in both short and long ISI conditions. This pre-training effect for short ISI suggests that speakers of a tone language have an advantage over speakers of a non-tone language in their "phonetic mode" of processing for F_0 (and possibly other perceptual cues to tone, such as harmonics). Just like the non-tone speakers, however, they cannot recruit the phonological level categories of the target language, and so show poor performance at the longer ISI pre-training. After training, the Chinese group showed an advantage over the English group in both the phonetic and phonological conditions (though still not native-like), suggesting that with experience, they had gained phonological knowledge which the English speakers had not. According to the authors, tone language speakers seem to have an advantage over non-tone speakers in learning a new tonal contrast in two ways: first, they can transfer their phonetic knowledge (F_0 tracking and other cues); second, they can more quickly apply phonological knowledge, either by mapping the new tones onto their native tone categories or by learning new ones, an ability which non-tone language speakers lack, because they have not yet acquired the requisite phonetic knowledge.

Francis et al. (2008) describe the explanation presented by Wayland and Guion (2004) as a *levels of representation* account, contrasting this with the *category assimilation* account of Hallé et al. (2004), which posits that both tone and non-tone language speakers process foreign tones in relation to their native categories, but that tone language speakers have more success than non-tone speakers because the latter's native categories are intonational, rather than tonal. The *levels of representation* account supposes that non-tone language speakers have no relevant phonological level for tones on which to draw, and therefore cannot use F_0 for lexical tasks; the *category assimilation* account supposes that non-tone language speakers do use phonological level categories, but because they are intonational, they are further away from the target categories than those of another tone language speaker — it does not suppose any differences in the ability to perceive the acoustic propeties relevant to tone.

Francis et al. (2008) also review findings by Wang, Behne, Jongman, and Sereno (2004), who found that Norwegian listeners do not show a left-hemisphere advantage in processing Mandarin tones, as do native Mandarin speakers, even though they are familiar with lexical tone categories in Norwegian (and for which they show a hemispheric effect), instead patterning with English speakers. Although Wang et al. (2004) conclude that this effect was due to Norwegian listeners' unfamiliarity with the specific perceptual correlates of Mandarin tone, it is worth noting that in order to equalize error rates across groups, Wang et al. (2004) manipulated signal-to-noise (S/N) ratio and ISI independently for each group based on pilot data; while the Norwegian subjects were grouped with English participants for S/N ratio, they were grouped with Chinese participants for ISI, possibly indicating some advantage in phonetic-level processing by the Norwegian speakers over the English speakers.

Francis et al. (2008) trained Mandarin and English speakers to identify Cantonese tones. Although both groups showed similar overall accuracy on pre- and post-training identification test, the results of the identification tests and a MDS analysis of pre- and post-training difference ratings revealed that, initially, each group perceived best those tones which could easily be mapped to their native categories. Mandarin listeners placed greater emphasis on *direction* than *height*, confusing tones which shared similar pitch contours, while English listeners placed greater emphasis on *height* than *direction*, confusing tones with similar average pitch height. Both groups improved after training, and had adjusted their weighting of the phonetic properties to more heavily favor average pitch height (which is an especially important dimension of tone in Cantonese; *cf.* Gandour (1979a, 1983)). The authors conclude that while native categories clearly influence how L2 categories are initially perceived, acoustic features also play a role when L2 categories cannot be assimilated (or when task demands, such as ISI, preclude access to phonological categories), and reweighting of phonetic properties is a key process during learning.

Another remarkable aspect of the finding by Francis et al. (2008) is the fact that Mandarin speakers in this study showed a stronger weighting for *direction* than *height* (in the pretest), unlike in previous studies (Gandour & Harshman, 1978; Gandour, 1983), where *height* was found to be the most prominent dimension for all language groups.

Differences in mismatch negativity (MMN) responses by Mandarin- and English-speaking listeners to Mandarin tones found by Chandrasekaran, Krishnan, and Gandour (2007b) are consistent with variation in sensitivity to dynamic properties of pitch. Mandarin-speaking listeners showed a greater MMN response to an acoustically dissimilar set of tones (Tone 1–Tone 3) compared to an acoustically similar set (Tone 2–Tone 3) in a passive oddball paradigm. English listeners did not show such a difference, and Mandarin and English speakers differed in their MMN response only in the high contrast condition. The authors speculate that the English listeners placed greater importance on pitch height, leading them to treat both tone sets as equivalently different, while the Mandarin listeners were influenced by their native language knowledge of the the dynamic properties of these tones (*i.e.*, direction and slope), and therefore perceived the tones as less similar to one another. A MDS analysis by Chandrasekaran, Gandour, and Krishnan (2007) confirmed this, revealing two dimensions, interpreted as *height* (linked to average F_0 and F_0 offset) and *contour* (linked to changes in the rate of F_0 change throughout the tone); Mandarin-speaking listeners relied more heavily on the *contour* dimension than did English-speaking listeners, and consistent with previous results (Gandour & Harshman, 1978; Gandour, 1983), *height* was important for both groups.

Wong and Perrachione (2007) trained English speakers to distinguish English pseudowords resynthesized with overlaid Mandarin lexical tones. Learners' pre-training ability to perceive the relevant pitch contours in a non-lexical (although still linguistic) pretest predicted their success of learning in the lexical task. Performance was also correlated with degree of musical training. These findings are interpreted as evidence for a "phonetic–phonological–lexical continuity," in which lower level knowledge acoustic knowledge must be established before being used for lexical tasks.

J. Lee et al. (2007) extended the results found by Wong and Perrachione (2007) by examining the effect of stimulus variability during training. Learners with higher pre-training pitch perception ability learned best with a high-variability training set (containing stimuli from four talkers), while those with poorer pre-training pitch perception learned best with a low-variability set (containing stimuli from one talker). Accomodating greater stimulus variability requires more robust phonological categories, and learners with less initial pitch perception ability have not yet refined the "phonetic categories [which] need to be established before the phonetic details are used phonologically, *i.e.*, to contrast word meanings". Thus, it appears that, while both groups improved in their ability to perform the lexical task, those with lower pitch perception ability were relying more on phonetic detail, while those with higher pitch perception ability more quickly learned and used categorical (phonological) knowledge. Chandrasekaran et al. (2010) further illuminated the nature of the advantage by some speakers, finding that better learners attend more to pitch direction than poorer learners, and that learning to use Mandarin tones to identify words increases the ability to identify pitch direction.

2.1.1.5.3 Acoustic

A key question about the results of Wong and Perrachione (2007), J. Lee et al. (2007), and Chandrasekaran et al. (2010) is to what degree the pre-training differences in pitch perception ability among their subjects are "general": first, are the learning differences observed attributable to a general pitch processing (or an even more general auditory) ability (acoustic), or specifically to the ability to use pitch in a speech context (phonetic)? Second, is this pitch ability applicable to pitch perception tasks generally, or just to the level, rising, and falling patterns examined?

Krishnan et al. (2005) found more accurate encoding of pitch in the auditory brainstem by Mandarin speakers compared to English speakers, as measured by the frequency following response (FFR). Specifically, the FFR of Mandarin speakers exhibited both stronger representation and smoother tracking of the fundamental frequency and second harmonic of Mandarin tones. Although the differences between the Mandarin and English speakers were specific to linguistically-relevant dimensions, rather than simply more accurate encoding of all aspects of the signal, these dimensions (F_0 and harmonics) could be relevant to a variety of speech and non-speech auditory tasks. Song, Skoe, Wong, and Kraus (2008) found similar effects on the brainstem pitch encoding of English speakers after only a short period of training on Mandarin tones, but only for Tone 3 (dipping), which is the most complex and unfamiliar tone for English learners.

Bent, Bradlow, and Wright (2006) demonstrated differences in the processing of non-speech sounds (pure tones) by English and Mandarin speakers. These differences were limited to the discrimination of pitch contours; there were no differences between the English and Mandarin groups on a nonspeech pitch discrimination task. In fact, the only differences between the groups were on certain falling and flat pitch contours, which Mandarin speakers misidentified more often than did English speakers. Based on a signal detection analysis, the authors suggest that this may have been due to response bias on the part of the Mandarin group (*i.e.*, they were treating the pure tones as speech sounds), rather than a difference in sensitivity. Nevertheless, this finding suggests either that the perceptual tuning to properties of pitch relevant to linguistic pitch representations is available for the processing of pitch information generally, or that nonspeech sounds are processed 'linguistically' if they share properties of linguistic pitch categories.

Y. Xu, Gandour, and Francis (2006) found that the differences between English and Mandarin speakers' perceptions of a continuum of level to rising speech tones were somewhat mirrored in their perception of non-speech sounds, which they theorize results from the interaction of domain general sensory memory and long term phonological memory. Importantly, while Mandarin speakers perceived both speech and non-speech stimuli in a similar categorical fashion, English speakers perceived the non-speech versions of the stimuli in a more categorical fashion than they did speech sounds. This could indicate that when English speakers perceive pitch in speech, they do not process it categorically because it does not match their stored representations of speech categories.

Similarly, Mattock, Molnar, Polka, and Burnham (2008) demonstrated that English-learning infants show a decrease in discrimination sensitivity to lexical tones, but not a corresponding decrease in sensitivity to similar nonspeech sounds. This leads the authors to argue for an early separation of speech and nonspeech perceptual processes. An alternative explanation is that that these results simply illustrate the early emergence of a "speech mode" of processing (Francis & Ciocca, 2003;

Burnham et al., 1996), and the first step in perceptual reorganization for infants is to ignore irrelevant dimensions in linguistic tasks, with loss of sensitivity to following later. Indeed, Francis and Ciocca (2003) provide evidence for such "speech mode" of processing, finding that Cantonese adult speakers showed an order effect in speech-tone discrimination, while English speakers did not; neither group showed such an effect for corresponding nonspeech tones.

Although the results of Y. Xu, Gandour, and Francis (2006) and Mattock et al. (2008) might suggest that language experience should not affect the perception of non-speech sounds, further studies illuminate the specificity of crosslinguistic differences and demonstrate that the language-induced tuning of brainstem FFR is not specific to speech sounds.

Y. Xu, Krishnan, and Gandour (2006) found that although English and Mandarin speakers did not differ in their FFR responses to linearly rising or falling nonspeech pitches, they did differ in response to nonlinear changes in nonspeech pitch, which more closely resemble the pitch patterns of Mandarin tones. Chandrasekaran, Krishnan, and Gandour (2007a) extended the results of Chandrasekaran, Krishnan, and Gandour (2007b) and Chandrasekaran, Gandour, and Krishnan (2007), finding that the MMN response of Mandarin speakers was greater than that of English speakers only for curvilinear pitch contours. This curvilinear feature could represent an additional, higher order perceptual dimension of pitch not previously identified in MDS studies using only linearly changing tones.

Krishnan, Swaminathan, and Gandour (2009) found that in response to Mandarin tones acoustically transformed to obscuring speech-specific acoustic information while preserving pitch (iterated rippled noise (IRN)), the FFR of Mandarin-speaking listeners exhibited smoother pitch tracking and more robust pitch representation; this was manifested not only in more robust representation of fundamental frequency, but in harmonics as well. Specifically, the FFR of Mandarin-speaking listeners more accurately encoded pitch information for *curvilnear*, as opposed to *linear* rising or falling stimuli, and represented spectral information up to the fifth harmonic, while that of English-speaking listeners represented spectral information only up to the third or fourth harmonic.

Krishnan et al. (2009) and Y. Xu, Krishnan, and Gandour (2006) note that the investigation by Y. Xu, Gandour, and Francis (2006) examined only linearly changing pitch contours, and that these linear approximations of the curvilinear Mandarin tones do not capture the language-specific tunings of the brainstem by language experience. According to the authors, these effects are ultimately domain-general, in that the auditory brainstem shapes multiple acoustic dimensions associated with pitch that are later processed by domain-specific mechanisms.

2.1.2 Summary

Lexical tone is the use of pitch in conjunction with segmental information to convey lexical meaning. Tone is primarily defined by F_0 patterns, which consist of several perceptual dimensions found crosslinguistically, three of the most important of which are *height*, based on the position of the tone within the reference range, *direction*, based on the direction of change in pitch over the syllable, and *slope*, based on the rate or shape of pitch changes.

The importance of these dimensions to the tonal system of a language varies based on its tonal inventory and tonal phonology, and linguistic experience with pitch causes a top-down tuning of pitch encoding in linguistically relevant ways which correspond to perceptual properties of pitch found in behavioral studies, and which in turn resemble theoretical features used to describe tonal systems. This tuning appears to extend to phonological, phonetic, and acoustic levels. This tuning, especially at the phonetic and acoustic levels, suggests a route by which linguistic experience may subsequently affect perception in other acoustic tasks, including music perception, which share a reliance upon the same acoustic properties.

2.1.3 Musical Pitch

If tones are the fundamental unit of pitch in language, then notes the fundamental units of pitch in music. Like a tone, a note is an entity in a rule-governed system, and can be given various names depending on its function with the system, such as *G*♯, or *sol*. This section will explain some acoustic facts relating to notes in music, how notes and note relationships are perceived, and how musical objects like melodies are built from individual notes, with particular emphasis on characteristics of musical pitch which parallel those found in language.

2.1.3.1 Acoustic and Musical Facts

The three basic properties of a musical note are its pitch, loudness, and timbre (Rasch & Plomp, 1982). Timbre is a catch-all term for those properties of a sound which impart its 'quality', for example, the difference between a trumpet and a saxophone playing the same note at the same loudness. Although timbre is important for music, it is generally considered secondary to pitch; in fact, (Patel, 2008b) argues that this is a key difference between language and music: linguistic categories are primarily timbral, while musical categories are primarily pitch-based. Lexical tone is an exception to the first generalization, but it remains that every language contains timbral categories (*e.g.*, consonants and vowels), while not every language contains tones.

Burns and Ward (1982) review the organization of frequency into musically meaningful units. Pitch categories in music consist of a finite set of notes, whose relationships to one another are defined by the ratio of their frequencies, termed the interval between the notes. The human auditory brainstem appears to be tuned to detect these frequency ratios (Tramo, Cariani, Delgutte, & Braida, 2003; Bidelman & Krishnan, 2009). Intervals with simpler frequency ratios (*e.g.*, 2:1, octave; 3:2, perfect fifth) are generally perceived as more pleasant, or consonant, than those with more complex (dissonant) ratios, which may be partially due to innate factors

(Zentner, 1998). Frequencies with a 2:1 ratio are perceived as having equivalent chroma (or note name identity; *e.g.*, 'B'), despite having a different pitch height, a phenomenon known as octave equivalence or octave generalization. Modern Western music divides the octave into twelve equal intervals, although only a subset of the twelve may be used in a given composition. This finite collection of pitch ratios within an octave range form a scale, which is the basis for the melodic and harmonic structure of music.

2.1.3.1.1 Melody

A melody is a collection of individual notes arranged linearly. It is typically composed from the notes of a particular scale, and contains an implied harmony, or harmonic progression suggested by the notes based on the scale and typical chord progressions of the musical culture and idiom.

Apart from its implied harmonic structure and rhythm, the content of a melody can be described in at least three ways: by its *key*, *contour*, and *intervals*. *Contour* encodes the up/down pattern of pitch changes, regardless of the note values, while *interval* encodes the precise distance from each note to the subsequent one. Here, *key* refers to the pitch range in which a melody occurs, or to what note it begins on; in most cases, melodies which differ only in key are considered to have the same musical identity.

Experimental evidence suggests that each of these are dissociable perceptual dimensions, with unique developmental courses (Trehub, Bull, & Thorpe, 1984; Trainor & Trehub, 1992; Trehub, Schellenberg, & Kamenetsky, 1999). Under most conditions, the *contour* of a melody is remembered more accurately by listeners than its *interval* content (Bartlett & Dowling, 1980; Edworthy, 1985; Peretz & Babaï, 1992; Trainor, Dejardins, & Rockel, 1999), and sensitivity to both *interval* and *contour* is enhanced by musical training (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004).

2.1.3.1.2 Musical Typology

The descriptions and generalizations given above apply to the Western musical tradition, which, due to increasing globalization, is pervasive in cultures around the world. The music of the various cultures of the world includes a great diversity of scales and harmonic systems besides that found in the Western (European) tradition, such as Indonesian gamelan (Vetter, 1989), and Indian ragas (Castellano, Bharucha, & Krumhansl, 1984); in addition, some styles of contemporary Western art music have developed tonal (or atonal) systems quite distinct from the roots of European folk music, such as 12-tone serialism.

Krumhansl (1990) reviews the results of several probe-tone studies on the perception of tonality in Western as well as several non-Western musical forms, including 12-tone serial, North Indian, and Balinese (gamelan) music. Listeners from each of these cultures (Western musicians well-versed in atonal music in the first case) were found to perceive the tonality and scale structure of the music in a manner consistent with music-theoretic descriptions of the relevant scale structure. Even non-enculturated listeners (*e.g.*, Westerners listening to gamelan music) could extract some information about pitch relations from the regularities in the non-Western music; however, they did not perceive pitch relationships that depend on knowledge of the underlying scale structure in a native-like way.

Thus, it appears that music is a culturally acquired system in a manner similar language (*i.e.*, according to innate predispositions which guide and shape the extraction of culturally-specific patterns from the environment (*cf.* Lynch & Eilers, 1991; Trehub et al., 1999; Trehub, 2003a, 2003b), although some purported universal properties of musical systems may allow even non-native listeners to extract some structure from an unfamiliar musical idiom.

Lynch and Eilers (1991) found that neither musician nor nonmusician (10–13-year–old) Western children could detect mistunings in the Javanese gamelan *pelog* scale, although both groups could detect mistunings in the Western major and minor scales, with musicians outperforming nonmusicians in this task. This study clearly demonstrates the effects of training at an early age, but another important finding is that even children who are not musically trained demonstrate a knowledge of the scales of their native musical forms. This task appears to draw less on perceptual ability than on knowledge of musical structure, and it appears that musicians trained in one musical culture cannot necessarily generalize this type of knowledge to another musical context.

Can the particular musical system acquired affect domain-general auditory perception? The available evidence suggests that sensory tuning to properties of the acoustic signal relevant to music drives auditory changes, similar to the effects of native language on the perceptual system.

Cooke (1992) examined the tuning tolerances of Ganda and Soga (Uganda) musicians, who are said to use an equi-pentatonic scale (five notes relatively equally spaced along the octave). Their tolerances for interval size were large when the intervals were presented sequentially; this is not unlike tuning judgments and production errors made by Western musicians. Although the tolerances of the Ugandan musicians were larger than those of the Westerners, this is to be expected based on the larger interval size of the pentatonic system of the Ugandans compared to the smaller intervals of the Western twelve note system; a pitch in a pentatonic scale can tolerate more variation before it intrudes on a neighboring pitch category.

A. Schneider (2001) describes the difficulties inherent in describing pitch systems, which result from the fact that the percept of musical pitch is more than simply fundamental frequency, although this is what Western music theory encodes. Many kinds of music (*e.g.*, gamelan) rely on instruments (gongs and metallophones) which produce sounds with inharmonic spectra and weak fundamentals, leading to ambiguity of pitch; these seem to be important features of gamelan music, which relies on interactions of these complex harmonics between instruments to create a "shimmering" perceptual effect, and may account for the wide variations in tuning found between different ensembles. A. Schneider (2001) speculates that "these components constantly interfere with the base [fundamental] frequency components and may be heard as a sequence with a melodic line of its own".

It appears that in music, just as in language, there are multiple cues to pitch, and the relative importance of these depends on the musical idiom, and it seems likely that the perceptions of listeners of a particular type of music will be influenced by its use of these cues.

2.1.3.2 Musicality

Another way of considering the effect of music on the perceptual system is the degree of musical competence, or musicality, obtained by the perceiver, and the majority of musical note perception experiments examine musicians alone, or in comparison to nonmusicians. Formal musical training includes practice not only at producing music, but at developing an ability to perceive fine auditory distinctions.

These highly developed motor and perceptual skills have a direct effect on the brains and behaviors of musicians (Münte, Altenmüller, & Jäncke, 2002; Besson, Chobert, & Marie, 2011). Musical training has been shown to affect higher cognitive functions such as auditory attention (Strait, Kraus, Parbery-Clark, & Ashley, 2010; Strait, Kraus, Abecassis, & Ashley, 2010), and temporal processing (Rammsayer & Altenmüller, 2006), and the brains of musicians show structural differences compared to nonmusicians, including more interhemispheric connections at the corpus callosum (Schlaug, Jäncke, Huang, Staiger, & Steinmetz, 1995; Schmithorst & Wilke, 2002) and increased gray matter in areas associated both auditory, motor, visual, and higher cognitive functions (Gaser & Schlaug, 2003).

Musicians devote more auditory cortex to the representation of piano (Pantev et al., 1998) and sinusoidal tones (P. Schneider et al., 2002), and their early cortical

ERPs to pure, piano, and violin tones display a larger amplitude than those of nonmusicians, correlating with degree of musical training (Shahin, Bosnyak, Trainor, & Roberts, 2003; Shahin, Roberts, & Trainor, 2004). Pantev, Roberts, Schulz, Engelien, and Ross (2001) found that musicians' neural responses were heightened for the timbre of their own instruments; that is, trumpeters responded more strongly to trumpet sounds than violin sounds, while violinists showed the opposite pattern. Even short-term training in melodic tasks results in plasticity and expansion of representation of scale tone in the auditory cortex (Pantev et al., 2003). The effect of music on perceptual ability thus appears to be one driven by the development of high-level cognitive skills as well as neuroplasticity of relevant properties of auditory sensory systems resulting from long-term exposure and training.

Although not denying a role for inheritance, Sloboda (1994) argues that becoming a skilled musician is more a matter of practice and experience than of innate ability or predisposition, and that the basic perceptual skills necessary for music are possessed in some form by nearly everyone, as members of a particular musical culture, with musicians developing them through training to a higher level. Although (Smith, 1997) reviews many studies illustrating fundamental differences between musicians and nonmusicians, he argues for the inclusion of nonmusicians in studies of music cognition, because music is an important part of the cultural and cognitive makeup, even for the unskilled listener, and a theory of music ought to describe this aspect of music as well. (Bigand & Poulin-Charronnat, 2006) review studies illustrating musical perception skills displayed by nonmusicians, who are described as "experienced listeners", despite a lack of explicit training. Nonmusicians behave similarly to musicians in many ways, and the authors note that for many experimentally demonstrated differences between musicians and nonmusicians, the effects are often simply larger or earlier for musicians, with the difference between musicians and nonmusicians often one of magnitude. (Bigand, 2003) argues that experienced listeners behave very similarly to musicians in regard to perception of musical structure, but that they may diverge in their sensory processing of musical sound, although this too may in some cases simply be a matter of degree.

A major difficulty in studying the effects of musical training is the particular definition of 'musician' used; for example, the musicians in studies by Alexander, Wong, and Bradlow (2005), Wong and Perrachione (2007), and Song et al. (2008) were those who had received six or more years of vocal or instrumental training beginning before age ten, while those in other studies (*e.g.* Schlaug, Jäncke, Huang, & Steinmetz, 1995) were defined as active professional musicians. Regardless of the definition used, musicianship appears to be a gradient ability, from the casual music fan to the conservatory student. Musical training affects musicians' perception of the sounds of music in several ways, including pitch, timbre, and timing (Kraus, Skoe, Parbery-Clark, & Ashley, 2009). Some pitch perception tasks tap into sensory properties of pitch (analogous to the *acoustic* or *phonetic* level in linguistic processing), while others drawn on "structural" properties of pitch, and the interrelatedness of pitch categories in the operant scale system (analogous to the *phonological* level).

2.1.3.2.1 "Phonological" Effects

In the phonological vein, Tervaniemi, Rytkönen, Schröger, Ilmoniemi, and Näätänen (2001) found that the MMN (both ERP and magnetoencehalography (MEG)) of musicians was sensitive to changes in the contour of a musical figure presented at different frequencies, reflecting invariance and generalization across keys. This is analogous to speaker normalization performed during linguistic processing. Subjects who could not accurately discriminate the deviant pattern, whether musicians and nonmusicians, did not show the effect, while the effect for those who could descriminate it grew over the course of the experience, reflecting short-term learning. The musicians in this study who most often displayed this effect were those whose experience was in genres which often do not use a score (*e.g.*, pop, jazz), rather than classical music, in which musicians typically perform from a score; thus, different kinds of musicality may rely more heavily on various modes of learning, and therefore shape perception differently.

Musicians display a greater MMN response to deviations in both melodic contour and interval structure compared to nonmusicians, but these groups show no difference in their response to frequency changes of pure tones (Pantev et al., 2003; Fujioka et al., 2004).

2.1.3.2.2 "Phonetic" and Acoustic Effects

At the level of acoustic detail, Micheyl, Delhommeau, Perrot, and Oxenham (2006) examined the pitch discrimination ability of professional musicians and non-musicians, finding a much smaller discrimination threshold for musicians, although nonmusicians could reach musician-like levels with only a few hours of training. The musicians' advantage held for both pure and harmonically complex tones, but was more pronounced for the complex stimuli. Musacchia, Sams, Skoe, and Kraus (2007) demonstrated more robust FFR representation of F_0 and and earlier onset of response to musical stimuli by musicians compared to nonmusicians. Pitch encoding was correlated with length of musical training, suggesting that the observed effects were due to musical training, rather than innate predispositions among the groups.

Bidelman and Krishnan (2009) examined interval perception using the FFR, finding that responses to dichotically-presented consonant intervals were more robust than those to dissonant intervals. The authors "infer that the basic pitch relationships governing music may be *rooted in low-level sensory processing* and that an encoding scheme that favors consonant pitch relationships may be one reason why such intervals are preferred behaviorally" (emphasis added). This suggests that such low-level architecture might be recruited by linguistic as well as musical pitch perception.

K. M. Lee et al. (2009) examined the brainstem encoding of vertical musical intervals among musicians and nonmusicians. For both consonant (major sixth) and dissonant (minor seventh) intervals, the FFR of musicians more robustly represents the harmonics (especially the second harmonic) of the upper tone of the interval compared to nonmusicians. According to the authors, the specificity of this effect for the upper tone reflects its compositional relevance and prominence in previous psychophysical and neurological studies; likewise, harmonics are particularly relevant for the detection of consonance and dissonance, and for the percep of timbre. The second harmonic above all others parallels the findings of Krishnan et al. (2005) for tone perception.

Some studies note an effect of musical training only on neural mechanisms for perception of abstract musical information, to the exclusion of basic sensory information (Pantev et al., 1998; Fujioka et al., 2004), others suggest that the effect of musicality may extend to more basic sensory processing (P. Schneider et al., 2002; Musacchia et al., 2007; K. M. Lee et al., 2009), at least for acoustic properties which are musically relevant.

A difference in behavioral performance, brain structure, or neural response between musicians and nonmusicians is not alone sufficient to conclude that musical training caused these changes, and not sufficient to reject the alternative hypothesis that those with prexisting advantages in pitch or rhythmic perception are more likely to pursue musical training; however, in many studies reviewed here, not only is there a group difference between musicians and nonmusicians, but these effects correlate with degree, length, or type of training within musicians (Tervaniemi et al., 2001; Shahin et al., 2003, 2004; Musacchia et al., 2007; Pantev et al., 1998, *inter alia*), suggesting that the effect is due to experience.

2.1.3.3 Summary

Taken together, the results of this body of research indicate that the human auditory system is tuned to perceive acoustic properties relevant to musical structure, even among listeners who are not formally trained musicians. Some of these properties (*e.g.*, F_0 height, pitch change over time, spectral composition) are relevant to both language and music, suggesting a possible link between the two.

2.2 Connecting Pitch in Music and Language

Kraus and Banai (2007) review research on how auditory processing changes in response to environmental and learning experiences. Many of these findings suggest that auditory processing malleability is controlled in large part by top-down (higher cognitive) functions. In regard to language experience, they review findings on the loss of sensitivity to non-native phonemes in infants, and enhancement of pitch processing and subcortical encoding among speakers of tone languages. Musicians' brains also respond more strongly to sound, and show the same subcortical enhancement as tone language speakers.

Changes in the auditory environment during development influences the structure of the auditory cortex as well, and cortical reorganization due to sensory loss is not uniform—it shows asymmetries particularly suited to regaining function (*e.g.*, visual cortex is used for hearing in the periphery by blind people). In regard to learning, cases where learning transfers from perception of one stimulus to another are particularly interesting; perceptual learning can take place as the result of several kinds of neural changes (*e.g.*, increased amplitude, increased precision, sharpened receptive fields, or reorganization of cortical maps). What is clear is that more learning takes place when the relevant characteristics of the stimulus are actively learned and attended to, compared to simple exposure.

Kraus and Banai (2007) cite Reverse Hierarchy Theory (RHT) (Ahissar & Hochstein, 2004; Ahissar et al., 2009) to account for the influence of top-down processes on perceptual learning, and one that might also explain transference of learning between domains such as language and music. RHT claims that neural changes begin at the highest cognitive level that can solve a task, with changes to lower areas following when needed. The tuning of cortical and subcortical pitch representation resulting from lexical training (Wang et al., 2003; Song et al., 2008) supports this theory, as do similar findings for vision demonstrating sensory tuning following visual categorization training (Jiang, Bradley, Rini, Zeffiro, & Vanmeter, 2006; Jiang et al., 2007).

The argument advanced by Kraus and Banai (2007) is different from other models, which suppose a greater degree of moduarity between language and music. Peretz, Champod, and Hyde (2003) argue that music-specific networks for pitch processing diverge early from those for linguistic pitch. The primary evidence for this is cases of amusia (tone-deafness) in which the processing of intonation pitch in apparenty unaffected. However, as reviewed above, the phonetic, linguistic, and processing characteristics of intonation are different than those of lexical tone, in ways which may be due to both qualitative differences and degree of complexity. Thus, this resut does not entirely rule out the existence of some shared networks for pitch between language and music.

The mechanisms of the RHT invoked by Kraus and Banai (2007) do not necessitate a total overlap between processing networks for musical and linguistic pitch, nor that crossover effects be entirely symmetrical. Instead, they predict that crossover effects occur in ways predictable from the content and context of learning.

Findings by (Strait, Kraus, Parbery-Clark, & Ashley, 2010) provide further support for the role of RHT in music–language crossover. Musicians and nonmusicians completed a variety of cognitive tests, including those presumed to be more (*e.g.*, backward masking) or less (*e.g.*, simultaneous masking) subject to cognitive control. Musicians outperformed nonmusicians on the backward masking perceptual task only, and their performance correlated with frequency discrimination ability. These abilities were uncorrelated for nonmusicians, suggesting that perceptual abilities among musicians are tuned via the influence of cognitive tasks which depend on sensory (auditory) input.

RHT could explain the parallels between language and music discussed by (Patel, 2008a) and others, which are united by certain abstract features; perceptual learning related to these features has the potential to cross domains, and even push down into lower level perceptions of both language and music. RHT is compatible with the Shared Sound Category Learning Mechanism Hypothesis (SSCLMH) articulated by Patel (2008a), which asserts that although language and music have very different structures, they share cognitive and neurobiological mechanisms of sound category learning—that is, language and music are two different systems of sound categories acquired by the same (or some of the same) perceptual processes.

The SSCLMH makes claims which contradict some models of language and music processing ike that of (Peretz et al., 2003) which suppose that the modularity of music and language extend to lower level auditory processing. However, the rincipes of SSCLMH are in accord with recent findings by Merrill et al. (2012) which trace the hierarchy of pitch representation in the brain for song and speech, finding a great deal of overlap for linguistic and musical representation in temporal cortical areas, with divergent areas for speech and music representation in frontal and parietal regions. This fits the predictions RHT and SSCLMH, in that more overlap is found in temporal "resource" networks encoding basic properties of sensory stimuli, and less overlap is found among regions higher in the hierarchy, which encode more abstract, task-specific "representations". This contradicts earlier views of acoustic processing (Zatorre, Belin, & Penhune, 2002) which claim that music and speech diverge earlier in the processing stream, but is consistent with developmental theories which link music and language learning to general perceptual processes and learning mechanisms (McMullen & Saffran, 2010; Trehub & Hannon, 2006).

The framework provided by RHT predicts that cross-domain effects on pitch perception should be found in both directions for properties which music and linguistic tone have in common. This matches recent findings demonstrating the influence of linguistic experience on music perception, and vice versa. Although these effects may be bidirectional in principle, based on shared dimensions of pitch, the contexts of learning for language and music are not the same. The OPERA hypothesis (Patel, 2011, 2012) enumerates some of these differences, and argues that, in practice, the effects of musical experience on speech perception may be greater than the reverse case. In addition to the *overlap* in neural resources for language and music, the training undergone by musicians is more *precise*, *repetitive*, *emotional*, and requires more deliberate *attention*, all of which enhance learning. For these reasons, the magnitude of effects from language experience to music, and from music experience to language may not be equal.

2.2.1 Musicality and Tone Perception

Several studies have found effects of musical background on tone perception tasks (Alexander et al., 2005; Delogu, Lampis, & Olivetti Belardinelli, 2006), word learning tasks involving tones (Wong & Perrachione, 2007), and neural responses to tones (Wong, Skoe, Russo, Dees, & Kraus, 2007).

Alexander et al. (2005) compared the tone indentification and discrimination abilities of American English–speaking musicians and nonmusicians, none of whom had any prior exposure to the Mandarin tones tested. The musicians were both faster and more accurate than the nonmusicians at identifying and discriminating syllables differing only in tone (though still not as accurate as native Mandarin speakers). This suggests that the musicians could transfer some part of their musical knowledge, which the authors suggest is pitch processing ability, to the linguistic task

Delogu et al. (2006) tested Italian speakers on a memory task requiring them to detect changes in the segmental or tonal content of a list of Mandarin syllables. All subjects were better at detecting segmental changes than tonal changes, but those subjects who demonstrated better melodic memory on a musical intelligence test (Wing, 1948) showed better performance compared to those with poorer melodic memory, but only for tonal detection. Although the authors claim to have demonstrated a causal relationship between melodic perceptual ability and lexical tone discrimination, the better tone discrimination and melodic memory could have been caused by other factors; the authors did not report the musical background of the participants, but this study suggests either that greater musical ability influences the processing of pitch for linguistic uses, or that a common aptitude underlies both tasks.

Wong, Skoe, et al. (2007) demonstrated how such transfer may take place. Recall that Krishnan et al. (2005) demonstrated more accurate encoding of pitch in the auditory brainstem by Mandarin speakers compared to English speakers, and that Song et al. (2008) demonstrated similar effects of learning for Englsh learners of Mandarin. Wong, Skoe, et al. (2007) compared the FFR in response to Mandarin tones of English speakers who had different levels of musical training. In addition to faster and more accurate perception of the tones, the FFR of musicians illustrated a more faithful neural representation of the F_0 contour; in particular, the accuracy of encoding of Tone 3, the most difficult tone for English listeners, correlated significantly with perception of Tone 3 and with years of musical training.

Musacchia et al. (2007) also demonstrated more robust FFR representation of speech F_0 by musicians (no language background was reported) for speech syllables (not based on Mandarin or another tonal language); musicians also demonstrated earlier onset of FFR responses. Strength of pitch encoding by musicians was correlated with length of training and amount of exposure to music, sugggesting that the differences observed between musicians and nonmusicians were due to musical training, and not to other underlying differences between the two groups.

These results suggest that musicians (who do not speak a tone language) can be said to resemble tone language speakers in some of the ways they perceive pitch; that is, musical experience improves the perception of dynamic properties of pitch in language.

2.2.2 Language Tonality and Music Perception

By contrast, less is known about how tone language speakers resemble musicians; that is, whether language experience enhances pitch perception in music.

Stevens, Keller, and Tyler (2004) found that Thai speakers were faster (but not more accurate) at detecting changes in the contour and intervals of two-note melody pairs.

Alexander, Bradlow, Ashley, and Wong (2011) found that Mandarin nonmusicians discriminated five-note melodies more accurately than English nonmusicians, but English listeners identified (matched with graphical representations) the melodies better than the Mandarin listeners.

Pfordresher and Brown (2009) found that tone language–speaking (Mandarin, Cantonese, and Vietnamese) nonmusicians more accurately discriminated and imitated two-note melodies than English-speaking nonmusicians, but there was no difference in absolute note errors. Likewise, the tone language group more accurately discriminated intervals but not individual notes compared to the nontone group.

2.2.3 Absolute Pitch

An important and much discussed language-to-music transfer effect concerns the prevalence of Absolute Pitch (AP) among different linguistic populations. AP is the ability to identify a specific note without a reference pitch.

A greater prevalence of AP among Mandarin-speaking musicians than among English-speaking musicians has been observed (Deutsch, Henthorn, & Dolson, 2004; Deutsch, Henthorn, Marvin, & Xu, 2006). The cause of this prevalence is unclear, because lexical tones are not defined by absolute pitch values, given the necessity of speaker and context normalization. Nonetheless, musicians with AP display differences in brain activity and hemispherization in areas also associated with language compared to non-AP musicians and nonmusicians (Schlaug, Jäncke, Huang, & Steinmetz, 1995), and in neural response to speech sounds (Oechslin, Meyer, & Jäncke, 2010), suggesting that there may be some connection between tone perception and AP.

Although AP was originally thought only to occur among musicians, Levitin (1994) found that when nonmusicians are asked to sing their favorite popular song, their production is very close to that of the original recording; Bergeson and Trehub (2002) made a similar finding for mother song. Levitin (1994) makes a distinction between absolute pitch memory, which does not necessarily require any special musical education or knowledge, and and absolute pitch labeling, which requires a specialized skill. Schellenberg and Trehub (2008) found that Chinese- and English-speaking children did not perform differently in an absolute pitch memory task which did not require labeling, which is consistent with this dissociation between recognition memory and labeling.

Because (pitch-labeling) AP is defined by the ability to assign labels to pitch categories, the prevalence of AP among Mandarin speakers could be due to domain-general improvements not in pitch perception or memory, but in the ability to assign labels to pitch categories. This is partially supported by Hsieh and Saberi (2008), who compared AP musicians trained in a fixed-*do* solfége system (one in which solfége syllables, like *do*, always correspond with the same note names; *e.g.*, C), with AP musicians trained in a movable-*do* system (one in which *do* is always the tonic note of the active scale, regardless of the note name). Hsieh and Saberi (2008) manipulated the pitch of solfége syllables to create a mismatch between their actual pitch and their intended value in a fixed-*do* system. Pitch identification by fixed-*do*–AP musicians suffered as a result of these manipulations, while that of moveable-*do*–AP musicians did not. This effect diminished as lexically relevant acoustic information was stripped away from the signal. This suggests that, at least for some AP musicians, there is a link between lexical memory and long-term memory for pitches. Although operating on a moveable-*do* system may be more similar to speaker-normalized tone perception, this result suggests the possibility that AP musicians and tone language speakers share mechanisms of storing pitch categories.

AP appears to be a complex ability which may be influenced by linguistic, as well as cultural, genetic, and musical factors (Levitin & Rogers, 2005; Dediu & Ladd, 2007; Dediu, 2010). In terms of the levels of representation framework, linguistic or ethnic effects on AP prevalence appear to be an effect at the "phonological" or "lexical" level or higher. Because lexical tone functions in a much different manner than absolute pitch reference, and because this ability is far from universal, even among musicians, it may be more relevant to the questions at hand to focus on the effects of language on relative pitch (although Hove, Sutherland, and Krumhansl (2010) present evidence for nonlinguistic ethnicity effects on relative pitch, as well).

2.2.4 Linking Melody and Tone Perception

The perception of both linguistic tone and musical melody relies on dynamic pitch information. Musical experience influences the perception of linguistic pitch, and speakers of tone languages perceive music differently than speakers of other languages; however, these effects have not yet been fully explained. Among the research questions asked by Kraus and Banai (2007) is "which acoustic elements of sounds are critical for language and music" in ways relevant to their mutual influence. Because tone language speakers are more sensitive than nontone language speakers to dynamic aspects of pitch in speech, then any effect of tone language experience on music perception is likely to be found in aspects of music which rely on dynamic pitch information. The following General Hypothesis summarizes this idea.

General Hypothesis Cases of influence of tone-language experience on melody perception, and vice-versa, result from domain-general auditory tuning in response to shared acoustic properties, from which are formed the important abstract structural properties of each domain; they are not the result of general changes in pitch perception beyond those required by the structural properties of the domain of experience.

Two properties of musical melody offer a potential homologue to the dynamic pitch properties important for tone perception. *Contour* encodes the up/down pattern of pitch changes in a melody, regardless of the note values, while *interval* encodes the distance from each note to the subsequent one. The contour and intervals of a melody are illustrated in Figure 2.1. Because *contour* encodes the direction of movement from one note to the next, without regard to the size of this movement, it is argued to correspond to the linguistic dynamic tonal property *direction* identified by Gandour and Harshman (1978) and others; *interval*, which describes the magnitude of change from one note to the next, is argued to correspond to the dynamic tonal property *slope*. Musical *key* encodes the level or range of a melody, independently of the internal content of the melody. *Key*, as used here, is a static property, and is argued to correspond to the static tonal dimension *height*. These melodic properties are illustrated in Figures 2.1 and 2.2, and the proposed mapping between melodic and linguistic pitch components is summarized in Table **??**.

It is important to note that this is a mapping between 'phonetic'-level properties, with the assumption being that similar phonetic properties are built from similar acoustic properties. Rather than a complete theory, the mapping proposed here is a starting point for investigating the relationships between these pitch properties, and properties may need to be added, removed, or their relationship revised upon the examination of data from the experiments presented in this dissertation.

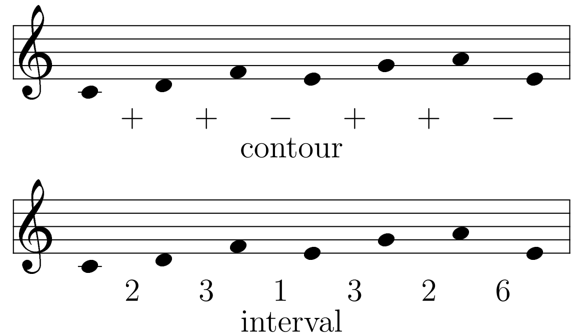

Figure 2.1: Contour and intervals for the same melody. + and − indicate rises and falls in pitch, respectively. Numbers indicate semitones.

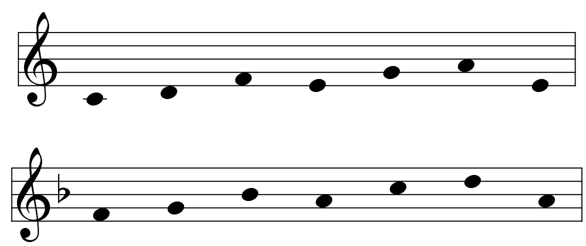

Figure 2.2: Two melodies with the same contour and intervals, but differing in *key*.

2.3 Remaining Questions and Plan of Experiments

The findings discussed in Section 2.2 indirectly support, but do not definitively establish the General Hypothesis or mapping between linguistic and musical components of pitch (Table **??**). The experiments described in Chapters 3–5 attempt to address these outstanding questions.

2.3.1 Language Tonality and Melody Perception

If corresponding properties rely on similar domain-general auditory mechanisms, then RHT predicts that crossover effects of linguistic or musical learning should be limited to those dimensions corresponding to those which are tuned, and should not affect other dimensions. This general hypothesis, that the effects of pitch experience on tone and melody perception are not across-the-board but are limited in scope, is consistent with many of the findings already discussed, such as those suggesting that the effects of musicianship on tone perception are driven by more accurate encoding of dynamic pitch information (Wong, Skoe, et al., 2007).

Effects in the other direction (language to melody) are somewhat more difficult to interpret within this framework. Stevens et al. (2004) drew an explicit parallel between the notion of 'contour' in music and language, finding that Thai speakers were faster than English speakers at detecting *contour* and *interval* changes in two-note melodies (transpositions/*key* changes were not tested). The authors claim that this indicates a specific effect, rather than general improvement of pitch perception, stating, "superior performance of the tonal language group on contour discrimination suggests that experience with a tonal language does not give rise to general pitch attunement but that it is specific to contour or pitch contrasts". However, because no differences were found between different kinds of melodic changes (*contour* vs. *interval*), this result does not rule out a general pitch attunement.

(Alexander et al., 2011) showed that Mandarin nonmusicians discriminate melodies more accurately than English nonmusicians, but the types of melodies used did not separately address *contour*, *interval*, and *key* changes.

(Pfordresher & Brown, 2009) found that tone language–speaking nonmusicians discriminate melodies and intervals better than English-speaking nonmusicians. The fact that no differences in individual note discrimination were found supports the idea that the language effects are not a general pitch attunement.

However, the relationship between *contour* and *interval* was not fully explicated in this study because of the short (two-note) stimuli used.

In order to address these questions, speakers of specific linguistic backgrounds must be compared on a carefully controlled musical task. Experiment 1 (Chapter 3) examines melody perception by speakers of three languages (English, Mandarin, and Yoruba) using the MET (Wallentin et al., 2010). Each language represents a distinct configuration of tonal properties (*i.e.*, nontonal, contour, register), leading to different predictions about their effects on perception of melodic *contour*, *interval*, and *key*.

2.3.2 Tone Learning and Melody Perception

Studies of the effect of lexical tones on melody perception have most often compared native speakers of tone and nontone languages. Although the human sensory and linguistic systems are not as plastic in adulthood, second language acquisition still causes perceptual changes in learners, including effects on pitch encoding in the auditory brainstem resembling those found among native tone language speakers and musicians (Song et al., 2008).

Experiment 2 and Experiment 3 (Chapter 4) attempt to compare the nature and magnitude of linguistic tone's effect on melody perception in adulthood by applying the MET to adult learners of Mandarin in cross-sectional (Experiment 2) and longitudinal (Experiment 3) designs.

2.3.3 Musical Experience and Tone Perception

Similarly to the converse case discussed above, studies of the effect of musicianship or musicality on linguistic have most often compared native speakers of tone and nontone languages; the effect of musical experience on tone perception has not been examined on a shorter time scale. Can short-term musical training lead to linguistic effects similar to those found after years of learning? The OPERA hypothesis predicts that this kind of training should be particularly effective at producing crossover effects.

Experiment 4 (Chapter 5) examines this question by examining changes in Mandarin tone perception by two groups of English-speaking musicians: one which undergoes aural skills training, and one which does not. This kind of musical ear training includes practice on melodic *contour* and *interval*, and so may affect the perception of tones relying on the analogous properties *direction* and *slope*, even in the absence of explicit linguistic training.

Chapter 3

NATIVE LANGUAGE TONALITY AND MELODY PERCEPTION

3.1 Introduction

The General Hypothesis formulated in Chapter 2 is that the effect of language tonality upon melody perception is not an across-the-board enhancement of pitch memory; rather, the effects are specific and attributable to properties of the tonal inventory of the speaker's language. Certain perceptual dimensions of pitch have been argued to underlie important structural elements of both lexical tone and melody, and it has been further argued that these dimensions are perceived by common general auditory mechanisms which are tuned by experience with linguistic and musical tasks. Based on the perceptual learning model provided of Reverse Hierarchy Theory (Ahissar et al., 2009), it was hypothesized that this tuning by experience with one kind of task (linguistic or musical) affects perception in the other when the auditory resources which are tuned are shared by the two tasks. The proposed mapping between perceptual dimensions of pitch in lexical tone and melody is summarized in Table **??**. This mapping generates hypotheses about how the relevant phonetic cues to tone in a given language will affect the perception of melody.

tone	**melody**
height	*key*
direction	*contour*
slope	*interval*

Hypothesis 1 Speakers of languages in which *direction* is an important phonetic cue to tone will demonstrate enhanced sensitivity to melodic *contour* compared to speakers of languages in which *direction* is not an important tonal cue.

Hypothesis 2 Speakers of languages in which *slope* is an important phonetic cue to tone will demonstrate enhanced sensitivity to melodic *interval* compared to speakers of languages in which *slope* is not an important tonal cue.

In this chapter, these hypotheses will be tested by comparing melody discrimination by native speakers of languages differing in the degree of importance of these dimensions. These languages were chosen because their tonal inventories differ according to the importance of these tonal dimensions. English is a non-tone language; as discussed in Section 2.1.1.3, speakers of non-tone languages appear to rely only on static properties like *height* when they attempt to perceive lexical tones. Mandarin is a contour tone language with four tones ((1)), for which *direction* and *slope* are important tonal cues (Gandour & Harshman, 1978; Gandour, 1983; Krishnan & Gandour, 2009, *inter alia*). Yoruba is a register tone language with three tones ((2)), for which *direction* is an important cue (Gandour & Harshman, 1978; Akinlabi, 2001, *inter alia*). Gandour and Harshman (1978) also suggested that *slope* is a somewhat important cue for Yoruba, based on the finding that Yoruba speakers are more sensitive to *slope* than are English speakers.

(1) Mandarin (Sino-Tibetan, China) (Wong & Perrachione, 2007)

a. mā (high-level, Tone 1) 'mother'

b. má (rising, Tone 2) 'hemp'

c. mǎ (dipping, Tone 3) 'horse'

d. mà (falling, Tone 4) 'scold'

(2) Yoruba (Niger-Congo, Nigeria) (Akinlabi, 2001)

a. kó (high) 'to build'

b. kō (mid) 'to sing'

c. kò (low) 'to reject'

The hypotheses enumerated above can now be specified in terms of these languages:

Hypothesis 1′ Speakers of Mandarin and Yoruba will demonstrate enhanced sensitivity to melodic *contour* compared to English speakers.

Hypothesis 2′a Speakers of Mandarin and Yoruba will demonstrate enhanced sensitivity to melodic *interval* compared to speakers of English.

Hypothesis 2′b Speakers of Mandarin will demonstrate enhanced sensitivity to melodic *interval* compared to speakers of Yoruba.

Hypothesis 2′b is somewhat speculative, as Mandarin and Yoruba speakers have not yet been compared directly in such a study, but Gandour and Harshman (1978) found that Thai speakers were more sensitive to *slope* than were Yoruba speakers, and suggested that this difference might arise from the lexical status of contour tones in Thai, while in Yoruba, contour tones do not have independent phonemic status. By extension, this difference may be expected to hold between Mandarin (like Thai, a contour tone language) and Yoruba, but this warrants further testing.

3.2 Experiment 1: Contour-tone, register-tone, and non-tone languages

3.2.1 Participants

Native speakers of English ($n = 20$), Mandarin ($n = 23$), and Yoruba ($n = 9$) participated in the study. Each participant completed a detailed language history questionnaire (Appendix B).

The English group consisted of monolingual speakers from the United States. Of the English speakers, 18 were female. The age of the English speakers ranged from 18 to 22 years ($m = 19.8$). Some English-speaking participants had studied other languages, but none had studied a tonal language, and none were native speakers of a language other than English.

The Mandarin speakers were all originally from China, but were currently working or studying in the United States; thus, they all had some level of proficiency in English, but none had native-like fluency. None had any experience learning another tone language, including significant exposure to non-Mandarin Chinese dialects (*e.g.*, Shanghainese). Of the Mandarin speakers, 11 were female. The age of the Mandarin speakers ranged from 19 to 34 years ($m = 24.3$).

Because English is the official language and the language of education in Nigeria, all of the Yoruba speakers who participated were also native speakers of (Nigerian) English. Of the Yoruba speakers, six were female. Some of the Yoruba subjects did not choose to provide their birthdate (all agreed to consent form certifying they were at least 18 years of age); the age of those participants reporting their age ranged from 19 to 26 years ($m = 26.9$).

Participants were excluded from the experiment if they had five or more years of experience studying a musical instrument or voice (including significant self-taught experience). Level of musicianship was assessed using a detailed music history questionnaire (Appendix B). Participants volunteered their time, or received a small course credit or cash payment in compensation.

3.2.2 Stimuli

Participants were administered a modified version of the Melodic Comparison subtest of the MET (Wallentin et al., 2010), a standardized test of music perception. the Melodic Comparison test is composed of melodies of 3–8 notes, each one measure in length and presented at a rate of 100 beats per minute (bpm). Melodies are presented in 52 pairs in a discrimination (AX) paradigm. Participants hear the two melodies, separated by a pause, and must indicate whether the two melodies are identical. In the original version of the MET, half of the melody pairs are "same" trials, and half are "different". In "different" melody pairs, a single note is changed from the first melody. In half of these "different" trials (13), the pitch of a single note is changed, but the *contour* of the melody remains the same; these trials will be referred to as *interval*-changing. In the remaining "different" trials, the note change results in a change to the *contour* of the melody; these trials will be referred to as *contour*-changing.

With the permission of its creators (Wallentin, p.c.), the Melodic Comparison subtest of the MET was modified to assess sensitivity to changes in the *key* of melodies. This was accomplished by duplicating and transforming the "same" trials from the original test, using the Audacity editor (Audacity Development Team, 2010). First, both melodies of the trial were transposed up or down by three semitones. Half of these transposed trials (13) were turned into "different" trials by transposing the second melody up or down by two semitones. These trials will be referred to as *key*-changing. The modified MET contains 78 trials (13 of each violation type, and an equal number of "same" trials), and lasts approximately 15 minutes.

3.2.3 Procedure

The modified MET was administered to participants via speakers. Participants indicated their response by checking a box on a paper form (Appendix B).

3.2.4 Results

The results of the MET were modeled as a signal detection task, in which the signal participants must detect is the change occurring in the second melody of a "different" trial (Macmillan & Creelman, 1991). A measure of sensitivity can be constructed based on the rate of hits (correct discrimination of a difference) and false alarms (discrimination of a difference where none exists). This is preferable to using accuracy as a dependent measure, because it takes into account each subject's overall bias to respond "same" or "different" on any trial. The measure of sensitivity used was A', which is the non-parametric equivalent of the commonly used d', chosen because it is appropriate for situations where some subjects score 100% or 0% in some condition, as occurred in these data (Pallier, 2002). A' scores range from zero to one, with $A' = 1$ indicating perfect discrimination and $A' = 0.5$ indicating chance performance. A' was calculated as described in Equation 3.1 for each change type (*contour*, *interval*, *key*) for each subject, and entered as the dependent variable in a mixed-effects analysis of variance with Violation Type (*interval*, *contour*, *key*) as a within-subjects factor, Subject L1 (Mandarin, Yoruba, English) as a between subjects factor, and Subject as a random factor. The results of the experiment are summarized in Table 3.1 and Figure 3.1.

$$A' = 1/2 + \frac{(hit - fa) * (1 + hit - fa)}{4hit(1 - fa)} \tag{3.1}$$

There were significant main effects of Violation Type ($F(2, 49) = 128.9245$, $p < 0.001$) and Subject L1 ($F(2, 49) = 5.2473$, $p < 0.01$). There was also a significant interaction effect between Subject L1 and Violation Type ($F(4, 49) = 4.1136$, $p < 0.01$). Post-hoc t-tests revealed that on the detection of *contour* changes, Mandarin speakers were more sensitive than English speakers ($p_{adj} = 0.043$). Yoruba speakers did not differ significantly from either Mandarin or English speakers in sensitivity to *contour*. On the detection of *interval* changes, Mandarin speakers ($p_{adj} = 0.003$) and Yoruba speakers (one-tailed $p_{adj} = 0.003$) were more sensitive than English speakers; Mandarin and Yoruba speakers did not differ significantly from one another in sensitivity to *interval*. None of the three language groups differed from one another in sensitivity to *key* changes.

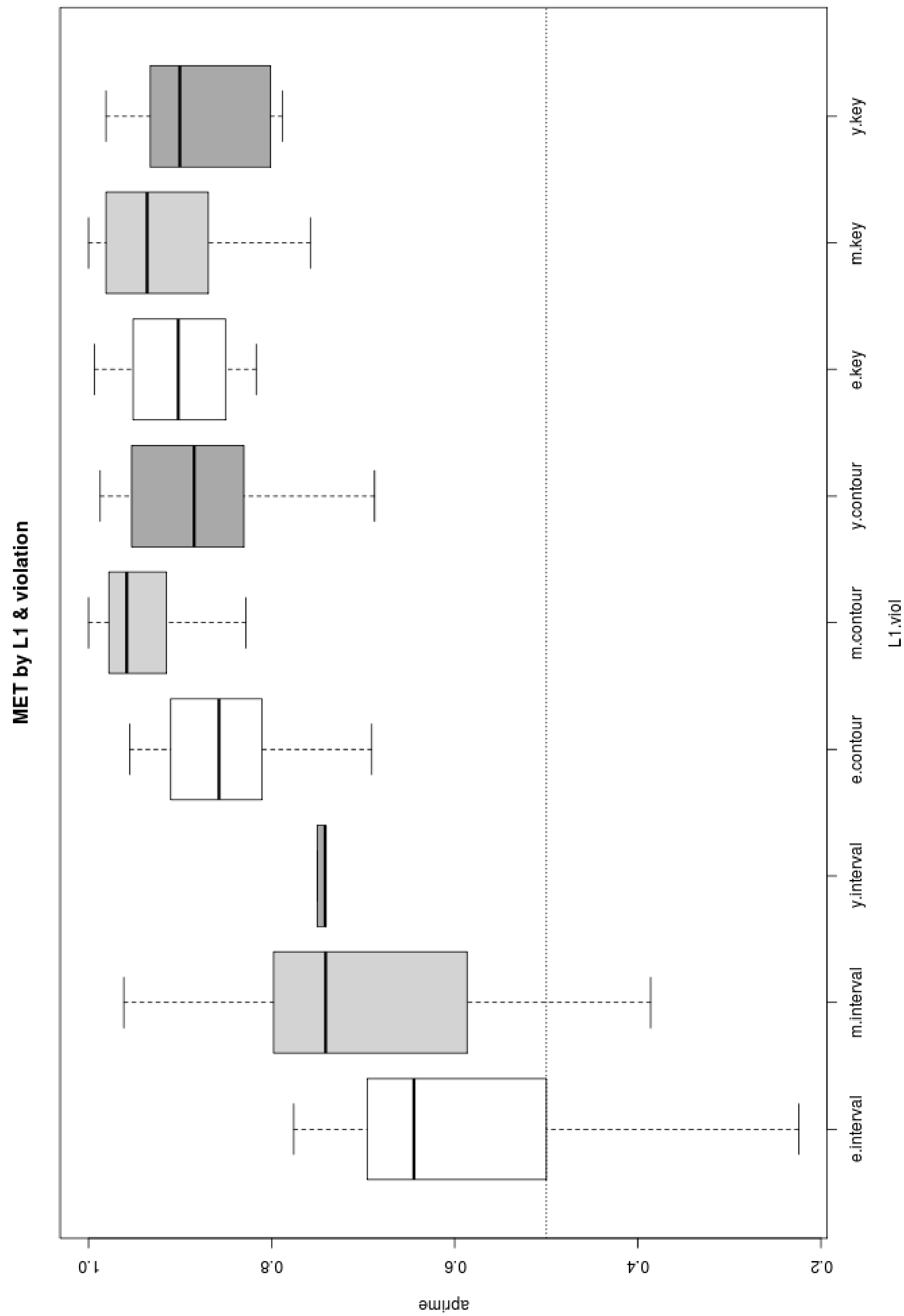

Figure 3.1: Sensitivity to MET violation types by language in Experiment 1. Dotted line indicates chance performance.

	interval	*contour*	*key*
English	0.588 (0.155)	0.852 (0.072)	0.898 (0.057)
Mandarin	0.699 (0.147)	0.936 (0.059)	0.919 (0.067)
Yoruba	0.732 (0.073)	0.878 (0.095)	0.856 (0.118)

Table 3.1: Mean A' scores (SD in parentheses) by violation type in Experiment 1.

3.2.5 Discussion

The general effects of violation type (with *interval* discriminated more poorly than the other violation types) is unsurprising, given what is already known about melody perception and the MET (Wallentin et al., 2010; Edworthy, 1985; Massaro, Kallman, & Kelly, 1980; Dowling, 1978, *inter alia*). The effects of language tonality on sensitivity to melodic structure matched the general prediction that differences in melodic discrimination between the three groups should not be the same for *contour*, *interval*, and *key*; some of the specific differences observed between language groups matched the effects predicted from the tonal inventories of the languages, while some differed from those predicted.

3.2.5.1 Mandarin vs. English

Mandarin speakers outperformed English speakers in the detection of *contour* and *interval* changes. Both of these differences were predicted (Hypothesis $1'$, Hypothesis $2'$a) based on the phonetic cues known to be relevant to tone perception in Mandarin. Importantly, Mandarin and English groups did not differ in sensitivity to *key* changes, indicating that the Mandarin speakers do not have better overall memory for melodies.

3.2.5.2 Yoruba vs. English

Yoruba speakers outperformed English speakers only on the detection of *interval* changes, and the two groups do not differ on *contour* or *key*. The difference on *interval* matches one hypothesized difference (Hypothesis $2'$a) between the groups.

Why did Yoruba and English speakers not differ on *contour* as hypothesized (Hypothesis 1′)? It is important to note that English speakers show fairly good discrimination of *contour* ($A' > 0.85$). This is not surprising given what is known about the melody perception generally (Wallentin et al., 2010; Massaro et al., 1980; Trehub et al., 1984), and could indicate the interference of ceiling effects in these data. The Yoruba group does not differ significantly from either the English or Mandarin groups, suggesting the possibility that there is a gradient effect among the three groups, which is not detectable due to the compression of scores in this sample. A replication with a more difficult test set is necessary to test this hypothesis.

3.2.5.3 Mandarin vs. Yoruba

Mandarin and Yoruba speakers were expected to differ in sensitivity to *interval* (Hypothesis 2′b), but did not. However, this hypothesis was highly speculative, and based on previous findings regarding Yoruba and Thai, which was perhaps unwarranted given the many differences among speakers of different contour tone languages in sensitivity to tonal cues (*cf.* Gandour, 1983, although only *direction* and *height* were considered). Because Mandarin and Yoruba speakers have not been compared directly on their sensitivity to *slope* in tone, it is possible that they do not differ in this way, and if so, they should not be expected to differ in sensitivity to *interval*. The assumptions made about Yoruba were based primarily on its tonal inventory, and the results of perceptual studies using tones in isolation. The perceptions of tones in continuous speech is more complex, and the perceptual cues relevant to register tone languages, including Yoruba, may be different than those following from the examination of the tonal inventory alone (*cf.* Yu, 2009). The direct comparison of Yoruba to the other tone language included in the several MDS studies discussed, as well as the extension of the MET to more tone languages, including Thai, would further illuminate this unresolved question.

There are a few alternative explanations for the observed results which bear consideration. First, *interval* perception is known to be difficult for nonmusicians (Massaro et al., 1980; Edworthy, 1985; Dowling, 1978). It is possible that although experience with *slope* influences the perception of *interval*, there is an upper limit to its influence on performance on the MET task, preventing the expected three-way distinction from emerging.

Alternatively, the proposed mapping between *interval* and *slope* may not be a perfect correspondance, and may require revision. Some demonstrated differences in pitch perception between Mandarin and English speakers are specific to curvilinear changes in pitch (Y. Xu, Krishnan, & Gandour, 2006; Chandrasekaran, Krishnan, & Gandour, 2007a; Krishnan & Gandour, 2009; Krishnan et al., 2010b), and the melodies in the MET contain discrete notes rather than continuous pitches. If the tonal property of *slope* describes only continuously changing pitch contours, then some other yet-to-be-identified tonal property relevant to Mandarin and Yoruba must be driving the effect on *interval*.

3.2.5.4 Bilingualism

The multilingual status of the Mandarin and Yoruba groups was not expected to be problematic for the hypotheses concerning tone language experience; no participants had any experience with a tone language other than their native language, and it has not been hypothesized that experience with a non-tone language alters sensitivity to the tonal cues of interest. However, these assumptions should be verified by comparing monolingual speakers of a nontonal language to speakers of two or more nontonal languages.

3.2.5.5 Mandarin dialects

The Mandarin-speaking participants in this experiment come from throughout China. Although all reported being native speakers of the standard dialect, there may still be considerable geographic variations in their speech, including in tones. To determine whether the previously discussed results might have been influenced by nonstandard varieties of Mandarin, Participants from Beijing, whose dialect most closely matches the standard, were compared to the other Mandarin-speaking participants.

MET scores were regressed on Violation Type (*interval*, *contour*, *key*) and location (Beijing or non-Beijing), with Subject as a random factor. The results revealed a significant main effects of violation type ($F(2, 20) = 61.3133$, $p < 0.001$), consistent with those found across languages, but no sgnificant difference between Beijing and non-Beijing Mandarin speakers. These results are summarized in Figure 3.2.

3.3 Summary

The results of the MET revealed that tone and nontone language speakers do not differ in sensitivity to every aspect of melodic structure. Speakers of English, Mandarin, and Yoruba did not differ in discrimination of *key* changes, but differ in discrimination of *contour* and *interval*, which are argued to correspond to tonal properties *direction* and *slope*. These properties are known to be less salient to speakers of nontone languages, so these results are consistent with part of the proposed mapping between structural properties of lexical tone and music (Table **??**).

Mandarin and Yoruba speakers did not differ from one another as expected, but did not pattern identically in comparison to English speakers; this suggests that the particular phonetic and phonological properties relevant to a language's tonal inventory may influence melody perception, but that the mapping may need to be refined, or the assumptions made about the languages in question revisited.

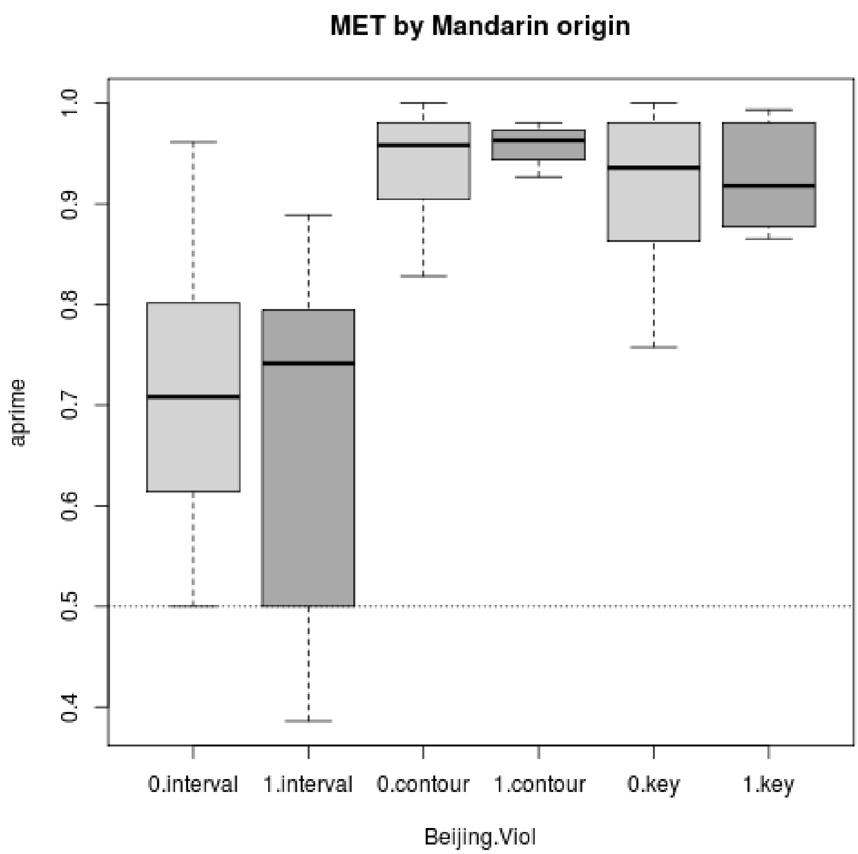

Figure 3.2: Sensitivity to MET violation types by Mandarin speakers from Beijing (0) and elsewhere (1) in Experiment 1. Dotted line indicates chance performance.

Chapter 4

SECOND LANGUAGE TONALITY AND MELODY PERCEPTION

4.1 Introduction

The results of Experiment 1 revealed that characteristics of the tonal inventory of a language have an effect on melody discrimination by speakers of that language, consistent with the General Hypothesis, and partially consistent with the proposed mapping between phonetic cues to tone and structural elements of melody summarized in Table **??**. The goal of the experiments presented in this chapter is to determine whether experience with a tone language in adulthood has qualitatively the same effects on music perception as observed in native tone language speakers.

tone	**melody**
height	*key*
direction	*contour*
slope	*interval*

Previous studies of lexical tone learning by non-tone language speakers have found changes in pitch processing after linguistic training (Wong & Perrachione, 2007; Wong, Perrachione, & Parrish, 2007, *inter alia*). The training paradigms used in these studies last multiple hours spanning several sessions, but are very short compared to the experience of native speakers. Nonetheless, such training induces changes in the encoding of pitch in the auditory brainstem (Song et al., 2008), which are similar to the neural differences between native Mandarin and English speakers (Krishnan et al., 2005) and represent the kind of tuning which has been suggested to underlie L1 effects on melody perception.

However, it is not known whether the tuning observed in response to such training is sufficient to lead to improvement in melody perception. Because the neural tuning observed in L2 tone learners is similar to that found in native speakers, if such crossover does occur, the General Hypothesis should be extended to include second language experience, and the effects of this experience on melody perception should be expected to resemble those observed in native speakers.

Hypothesis 3: L2 learners of a tone language will show enhanced discrimination of musical melody in a manner similar to that of native speakers of the language.

However, the time course of the emergence such an effect is still unknown; nor is it known what level of tonal proficiency is neccesary to induce such an effect. Hypothesis 3 will be tested in two experiments involving native English speakers learning Mandarin. Experiment 2 examines the change in melody discrimination in beginning Mandarin learners over time, and Experiment 3 compares learners with different levels of experience learning the tone language.

4.2 Experiment 2: L2 Mandarin study and melody discrimination

Experiment 2 attempts to examine the effect of Mandarin experience on melody perception in learners over time. Hypothesis 3 can now be further specified in terms of the language of interest:

Hypothesis 3′: L2 learners of Mandarin will show enhanced discrimination of melodic *contour* and *interval*, relative to L2 learners of a non-tone language.

4.2.1 Participants

Students enrolled in two sections of an introductory Mandarin language course at the University of Delaware took part in the experiment. These students were studying Mandarin for the first time. Sixteen students in the Mandarin group began the study, but only 10 students completed both phases of the experiment; only data from those participants who completed the study will be considered. Of these, five were female. The age of the participants ranged from 18 to 21 years ($m = 18.5$).

A control group of students in an introductory German language course also participated. These students were studying a second language, but this language is not tonal, and so no improvement in melody perception was predicted under the current hypotheses. Ten students in the German group began the study, but only eight students completed both phases of the experiment; only data from those participants who completed the study will be considered. Of these, three were female. The age of the participants ranged from 18 to 20 years ($m = 18.8$).

Both the Mandarin and German classes met five days per week during the academic term for 50 minutes. Participants from both groups were native speakers of American English, and had no prior experience with a tone language, and none more than five years experience studying a musical instrument or voice. Language background and level of musicianship were assessed using detailed language and music history questionnaires (Appendix B). Participants received a small course credit for participating.

4.2.2 Procedure

The modified melodic comparison subtest of the MET, described in Section 3.2.2, was administered to participants at two time points. The first administration (henceforth, the "pretest") was administered during the first two weeks of the academic term, and the second (henceforth, the "posttest"), was administered near the end of the same term, between 75 and 86 days later ($m = 82.8$).

4.2.3 Results

A' was calculated as described in Equation 3.1 for each change type (*contour*, *interval*, *key*) for each subject. Scores were regressed on L2 (German or Mandarin), session, and violation, with session and subject as random effects. The results of this model are summarized in Table 4.1. MET scores are described in Table 4.2 and summarized graphically in Figures 4.1–4.4.

There were significant differences in sensitivity to each violation type; *contour* and *key* were discriminated better than *interval*, which is consistent with other experiments reported here and elsewhere. There was no significant change in performance, either on overall MET score, or for the individual violation types, for either the Mandarin or German groups.

$$A' = 1/2 + \frac{(hit - fa) * (1 + hit - fa)}{4hit(1 - fa)} \tag{3.1}$$

	β	Std. Error	t	p
(Intercept)	0.64238	0.08171	7.862	
L2[Mandarin]	-0.09175	0.10963	-0.837	
violation[contour]	0.27191	0.10743	2.531	0.022
violation[key]	0.26869	0.10743	2.501	0.024
session	0.03166	0.04956	0.639	
L2[Mandarin]:violation[contour]	0.03308	0.14414	0.230	
L2[Mandarin]:violation[key]	0.03771	0.14414	0.262	
L2[Mandarin]:session	-0.04379	0.06649	-0.659	
violation[contour]:session	-0.06982	0.06795	-1.028	
violation[key]:session	-0.03809	0.06795	-0.561	
L2[Mandarin]:violation[contour]:session	0.03889	0.09116	0.427	
L2[Mandarin]:violation[key]:session	0.06191	0.09116	0.679	

Table 4.1: Estimates of fixed effects on MET scores in Experiment 2.

	interval	
	pretest	posttest
German	0.692 (0.086)	0.692 (0.085)
Mandarin	0.530 (0.212)	0.543 (0.206)
	contour	
	pretest	posttest
German	0.864 (0.071)	0.864 (0.077)
Mandarin	0.803 (0.073)	0.796 (0.087)
	key	
	pretest	posttest
German	0.914 (0.043)	0.906 (0.068)
Mandarin	0.891 (0.103)	0.876 (0.126)

Table 4.2: Mean A' scores (SD in parentheses) for the MET in Experiment 2.

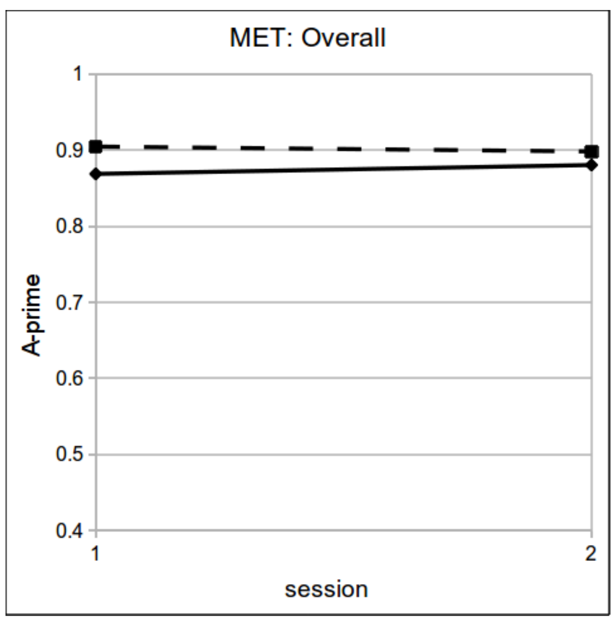

Figure 4.1: Overall MET performance by group (solid=Mandarin, dashed=German) in Experiment 2.

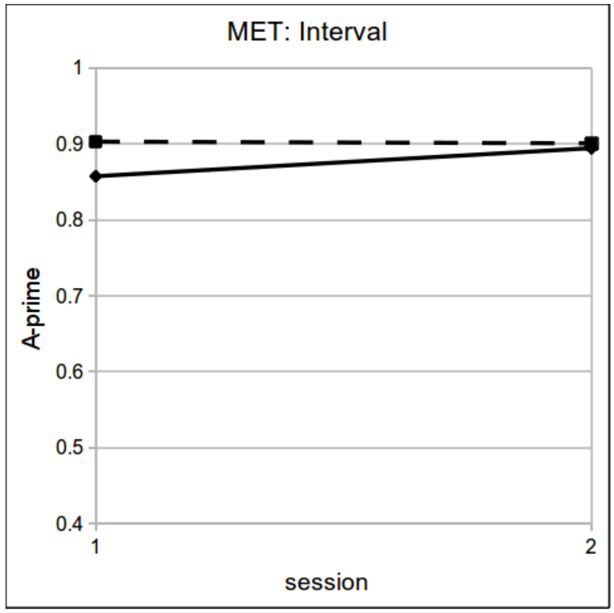

Figure 4.2: MET *interval* sensitivity by group (solid=Mandarin, dashed=German) in Experiment 2.

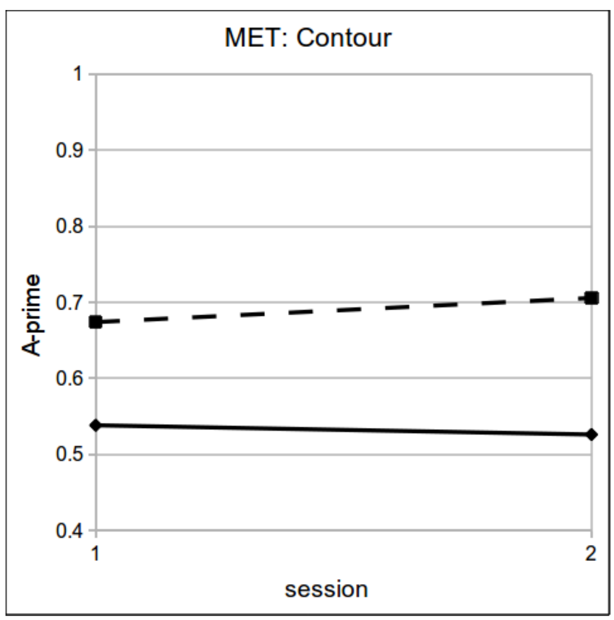

Figure 4.3: MET *contour* sensitivity by group (solid=Mandarin, dashed=German) in Experiment 2.

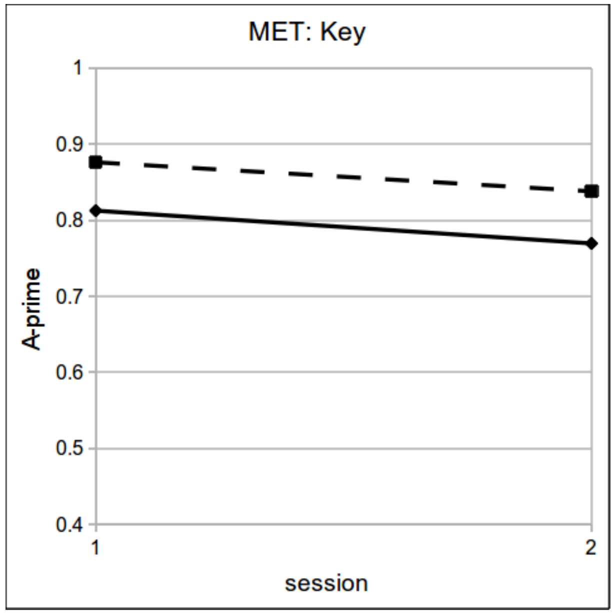

Figure 4.4: MET *key* sensitivity group (solid=Mandarin, dashed=German) in Experiment 2.

4.2.4 Discussion

No reliable improvement in melody discrimination was observed between pretest and posttest among English-speaking learners of either Mandarin or German. The primary limitation of this study was the small sample (see Section 4.3.5), but there are several possible explanations for this unexpected result.

First, although some studies (Song et al., 2008, *inter alia*) have indicated that subcortical encoding of pitch occurs after only a few weeks of training, these studies used training paradigms focused only on word-learing or tone discrimination, rather than comprehensive language learning; it is possible that such changes occur at a slower rate when learners must spread their effort between learning tones, words, syntax, and other aspects of the language, and that the learners examined here had not reached a sufficient level of phonetic proficiency to register an effect. The addition of a tone perception test, and the application of a tone-focused learning paradigm, rather than comprehensive language learning, would provide a better comparison to previous research.

Second, even if the reported differences in linguistic pitch encoding did emerge in the naturalistic setting examined over the time period observed, their effect on melody perception may not occur immediately, but may follow at some delay. Referencesexp:L2b will attempt to probe the L2 tone–melody relationship further by examining Mandarin learners with more tone language experience than those examined in Experiment 2.

4.3 Experiment 3: Length of L2 Mandarin study and melody discrimination

The results of Referencesexp:L2a, in which no difference was found in MET improvement over time between students learning Mandarin and those learning German, was attributed to the relative lack of proficiency acquired after only one term of study of Mandarin.

In order to determine whether second language experience with lexical tones leads to effects on melody perception comparable to those found among native speakers, and what the time course of the emergence of these effects may be, the current experiment examines melody perception among groups of L2 Mandarin learners with different levels of proficiency.

Hypothesis 3′b: L2 learners of Mandarin studying at a higher level will show enhanced discrimination of melodic *contour* and *interval*, relative to learners of a non-tone language.

4.3.1 Participants

Students enrolled in three Mandarin language courses participated in the experiment. These three courses represent the first three courses in the Mandarin language curriculum at the University of Delaware. These classes will be referred to as Levels 1, 2, and 3, with Level 1 being the lowest (introductory) level class.

Seven students from Level 1 (four female), six from Level 2 (three female), and eight from Level 3 (two female) participated in the study. The age of the participants ranged from 18 to 26 years ($m = 20.0$).

All participants were native English speakers, and had no prior experience studying a tone language other than Mandarin. None had more than five years experience studying a musical instrument or voice. Language background and level of musicianship were assessed using detailed language and music history questionnaires (Appendix B).

4.3.2 Procedure

The modified melodic comparison subtest of the MET, described in Section 3.2.2, was administered to all participants at the beginning of the academic term; thus, students in the Level 1 class had not yet begun their Mandarin study, while those in the Level 2 and Level 3 classes had completed one and two semesters, respectively.

4.3.3 Results

A' was calculated as described in Equation 3.1 for each change type (*contour*, *interval*, *key*) for each subject. Scores were regressed on L2 (German or Mandarin), session, and violation, with session and subject as random effects. The results of this model are summarized in Table 4.3. MET scores are described in Table 4.4 and summarized graphically in Figure 4.5.

There was no significant effect of class level; the three groups did not differ from one another, and there was no trend of increasing MET performance as class level increased, either for overall MET score or for individual violation types.

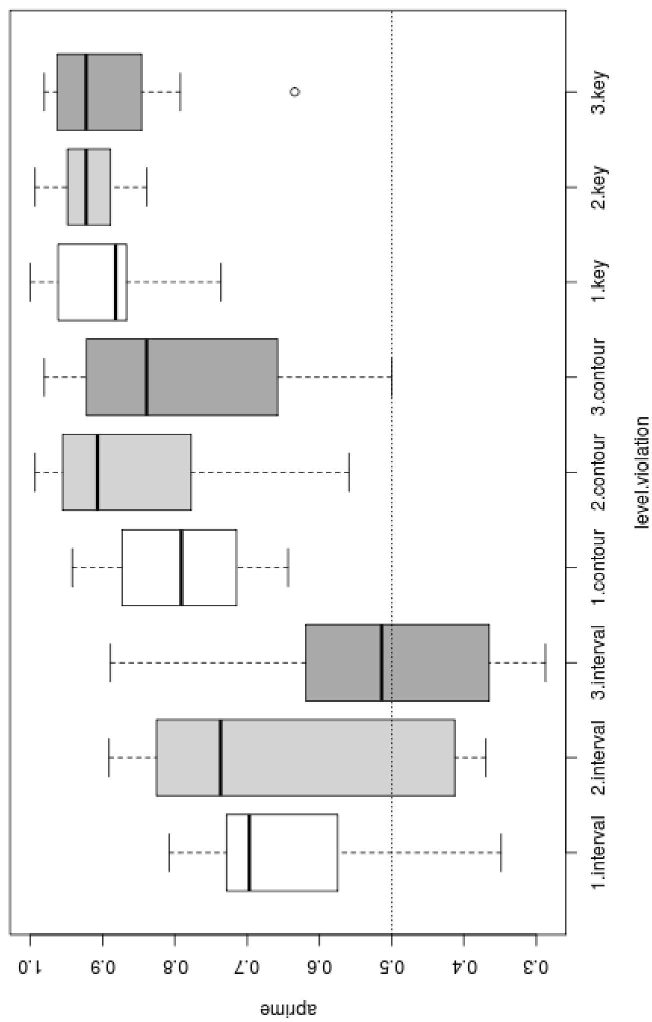

Figure 4.5: Overall MET performance by class level in Experiment 3.

	β	Std. Error	t	p
(Intercept)	0.72162	0.08540	8.450	
level	-0.05930	0.03856	-1.538	
violation[contour]	0.09199	0.06914	1.330	
violation[key]	0.19007	0.06914	2.749	0.025
level:violation[contour]	0.05662	0.03122	1.813	
level:violation[key]	0.05288	0.03122	1.694	

Table 4.3: Estimates of fixed effects on MET scores in Experiment 3.

	interval	*contour*	*key*
Level 1	0.637 (0.153)	0.793 (0.109)	0.896 (0.088)
Level 2	0.662 (0.220)	0.850 (0.162)	0.919 (0.053)
Level 3	0.522 (0.198)	0.790 (0.171)	0.885 (0.117)

Table 4.4: Mean A' scores (SD in parentheses) for the MET in Experiment 3.

4.3.4 Discussion

No reliable improvement in melody discrimination was observed across the three levels of Mandarin-learning students observed, and the small sample size renders interpretation or any trends difficult (see Section 4.3.5). As discussed in response to Experiment 2, the level of tone language proficiency required to cause changes in music perception is unknown, and class level is a very crude stand-in for proficiency. Moreover, participants in the highest class examined (Level 3) can still only best be described as 'intermediate' learners, and so the Mandarin learners here represent only a small portion of the continuum from beginning to fluent Mandarin proficiency. The addition of a tone perception test, as well as replication with advanced and near-fluent learners of Mandarin, would provide additional perspective on the effects of L2 tone proficiency, if any, on music perception.

4.3.5 Power in Experiment 2 and Experiment 3

A power analysis was conducted to determine to what degree the sample size affected the ability to detect an effect in Experiment 2 and Experiment 3. The results

	N	d
exp 1 (interval)	20, 23	0.735
exp 1 (contour)	20, 23	1.282
exp 2–3 (interval)	30	0.735
exp 2–3 (contour)	11	1.282
exp 2–3 (interval)	117	0.368
exp 2–3 (contour)	39	0.641

Table 4.5: Power analysis summary for Experiment 2 and Experiment 3. Sample (N) and effect (d) sizes in Experiment 1, and estimated sample sizes to detect equivalent and smaller (half as large) effects in Experiment 2 and Experiment 3 for power level $\pi = .8$ and significance level $\alpha = .05$.

of this analysis are summarized in Tablẽeftab:power. The effect size (Cohen's d) was calculated for *interval* and *contour* effects in Experiment 1. In order to detect an equivalent effect size—that is, a difference of similar magnitude in mean A' scores between Mandarin and German learners in Experiment 2, or between the lowest and highest level Mandarin leaners in Experiment 3—a substantially greater sample is needed. If the effect size in second language learners is smaller, as is lkely with their relatively low proficiency, an even greater sample is necessary.

4.4 Summary

The results of Experiment 2 and Experiment 3 fail to support Hypothesis 3, the extension of the General Hypothesis to L2 contexts, or the specific effects predicted in Hypothesis 3′b. It may be that only tone language experience during a critical period (native experience) is sufficient to cause such effects, either because the effects of on the linguistic system in first and second language contexts are different, or because the cognitive and neural resources underlying the crossover between language and music are less active in adulthood. Although these hypotheses were not supported, due to the limited sample, further studies including participants with a wider range of L2 tone proficiency, along with the use of more detailed measures of this proficiency should be conducted before completely abandoning Hypothesis 3.

Experiment 4 further examines the effects of experience with pitch learning over a short-term interval in adulthood, but from the converse perspective of musical learning on L2 tone perception.

Chapter 5

EFFECTS OF AURAL SKILLS TRAINING ON LEXICAL TONE PERCEPTION BY MUSICIANS

5.1 Introduction

The results of Experiment 1 suggests that perceptual tuning to phonetic cues to tone leads to greater sensitivity to corresponding dimensions of melody, as summarized in Table **??**. The reverse case, that musical experience leads to enhancement of lexical tone perception by nontone language speakers, has already been established (Wong, Skoe, et al., 2007, *inter alia*), and the observed effects can be generally characterized as better tracking of pitch movement on the part of musicians (Chandrasekaran, Gandour, & Krishnan, 2007), which is consistent with the General Hypothesis and proposed mapping between properties of tone and melody (Table **??**).

tone	**melody**
height	*key*
direction	*contour*
slope	*interval*

The context of experience considered in previous studies (including some experiments reported in this work) is that of years of musical training and practice, and makes little distinction between instrumental or vocal music lessons and other kinds of musical training. In addition to performance practice, professionally trained musicians often study music theory, including aural skills or "ear training". This kind of training focuses on learning to recognize and produce important elements of musical structure, including melodies and melodic intervals. While normal musical performance requires the ability to perceive melodic structure, such explicit perceptual training is expected to enhance the perception of lexical tone beyond mere performance of music. The OPERA hypothesis (Patel, 2011, 2012) predicts that this intensive, explicit kind of training specifically enhances plasticity.

Hypothesis 4: Nontone language speakers receiving aural skills training will show improvement in discrimination of lexical tones compared to those not receiving such training.

5.2 Experiment 4: Musical ear training and lexical tone perception

In order to examine the effect of aural skills training on the perception of linguistic pitch, the Mandarin tone discrimination ability of English speakers who already possess a high level of musicality was examined before and after a course of ear training. Those receiving training are expected to show a change in sensitivity to lexical tones after musical training.

Hypothesis 4′: English-speaking musicians who receive aural skills training will display improved performance in Mandarin lexical tone discrimination relative to musicians of approximately comparable ability who do not receive such training.

5.2.1 Participants

Participants were drawn from three groups: college music majors undergoing ear training (henceforth, 'aural skills musicians'), individuals with music performance experience who lack explicit perceptual training in music (henceforth, 'amateur musicians'), and those without music performance experience (henceforth, 'nonmusicians').

The music majors in the aural skills musician group have a great deal of music performance experience, but they are just beginning their first advanced study of music theory and aural skills. The amateur musician group was chosen for comparison because they were expected to possess approximately the same level of experience and musicality as the aural skills musician group, but would not be undergoing formal music theory training. The nonmusicians have less experience that the two musician groups, and do not participate regularly in musical activities; they provide an additional estimate of the re-test effects for the musical and linguistic measures used.

The aural skills musician group consists of freshman students enrolled in their first semester of aural skills training. These students were majoring in music at the college level and had studied their primary instrument or voice for more than five years. Fifty-one aural skills musicians began the study, but only 30 completed both phases of the experiment; only data from those participants who completed the study will be considered. Of these, 21 were female, and the age of the participants ranged from 18 to 20 years ($m = 18.5$).

The amateur group consisted of individuals who were not majoring in music, nor had they studied music theory or aural skills at the college level. They had studied an instrument or voice for more than five years and actively participated in a musical ensemble. Fourteen amateur musicians began the study, but only 13 completed both phases of the experiment; only data from those participants who completed the study will be considered. Of these, six were female. The age of the participants ranged from 18 to 21 years ($m = 19.4$).

The nonmusician group consists of individuals who had studied a musical instrument or voice for fewer than five years, and do not actively participate in music performance activities. Eleven nonmusicians began the study, but only 9 completed both phases of the experiment; only data from those participants who completed the study will be considered. All of these participants were female and their age ranged from 18 to 21 years ($m = 19.4$).

All participants were native speakers of American English and had never studied a tone language. Language background and level of musicianship were assessed using detailed language and music history questionnaires (Appendix B). Participants volunteered their time or received a small course credit or cash payment for participating.

5.2.2 Stimuli

Stimuli consisted of Mandarin words and pseudowords recorded by a female native speaker. The words and pseudowords were monosyllables spoken with each of the tones of Mandarin, as illustrated in Table 5.1. With the exception of minor phonetic details, their segmental content is licit in English; thus, the only 'foreign' content in the words is tone. Some of these words correspond to real words in Mandarin, while some are pseudowords; however, this distinction is meaningless to the participants, who have no knowledge of any Mandarin words.

	ku	*di*	*ma*
Tone 1	kū	dī	mā
Tone 2	kú	dí	má
Tone 3	kŭ	dĭ	mă
Tone 4	kù	dì	mà

Table 5.1: Mandarin (pseudo)word set used in Experiment 4.

In order to prevent non-pitch correlates of tone from influencing performance, the words were acoustically manipulated using the Pitch-Synchronous Overlap and Add (PSOLA) method in Praat (Boersma & Weenink, 2010), with the following process:

1. Two tokens of each word were spoken by the speaker within the carrier phrase in (1); this phrase was selected to avoid interference from tone sandhi.

2. For each of the three syllables, this resulted in two sets of four words differing only in tone. Each set of words was processed using the following steps:

   (a) The length of each token was normalized to the average length of the four words in the set.

   (b) F0 information was extracted from each of the four utterances of that syllable.

   (c) These pitch tracks were overlaid (using the PSOLA method) onto the segmental content of a Tone 1 token to create four new tokens differing only in pitch content.

(1) Shūo “______” de nà -ge rén lái -le.
say ______ REL that -CL people come -ASP
‘That person who said “______” came.’

The resulting Mandarin stimuli were compiled into a Mandarin Word Discrimination Test. Trials were constructed by presenting words in pairs, separated by a 500 millisecond pause. Within each trial, the two words always had the same segmental content. Each tonal contrast (*e.g.*, Tone 1 vs. Tone 2) was tested 12 times in total: twice in each order for each of the three syllables. An equal number of ‘same’ trials were included; all ‘same’ trials were composed of two different tokens of the same word. The test contained a total of 144 trials and lasted approximately 10 minutes. A full list of stimuli and trials can be found in Appendix A.

The stimuli were verified by two native speakers of Mandarin who did not participate in constructing the stimuli.

The native speakers heard each of the 24 stimuli in pseudorandom order, and were asked to identify each word by writing it down. They were also asked to rate the naturalness of each word on a five-point scale, with 1 indicating a low degree of naturalness, and 5 indicating a completely natural word.

	1	2	3	4
1	-			
2	1	-		
3	1	0.9545	-	
4	1	1	1	-

Table 5.2: Sensitivity (A' scores) by contrasts in the Mandarin Tone Discrimination Test by native Mandarin speakers.

The tones of all stimulus words were identified correctly by both listeners. One listener misheard two 'di' stimuli as 'bi', but this is not expected to influence the discrimination test, as participants in that test are informed that each word pair differs only in tone, if at all.

Naturalness ratings indicate that native speakers perceived the stimuli as natural. Most stimuli were rated as completely natural (5), and all but one were rated 4 or above. Identification scores and naturalness ratings for each stimulus are included in Appendix A.

The same speakers took the discrimination test. The results were analyzed in the same manner as for the English-speaking participants, described in Section 5.2.4.2. The results are summarized in Table 5.2. The native listeners displayed perfect discrimination on every tone contrast except Tone 2–Tone 3, although they A' was still over 0.95. This indicates that this contrast is non-trivial even for native speakers.

5.2.3 Procedure

The experiment was conducted in two sessions. For the aural skills musician group, the first session (henceforth, the "pretest") was administered during the first two weeks of the academic term, and the second (henceforth, the "posttest"), was administered near the end of the same term, 77 days later. Participants in the other groups completed the pretest and posttest between 20 and 191 days apart

($m = 62.2$).

5.2.3.1 Mandarin Word Discrimination

Participants were told that they would be listening to words from a foreign language, and that in this language, the pitch pattern of a word is a determining factor in its meaning, in addition to the vowels and consonants. They were then presented with examples of each of the four tones of the language, and completed two practice discrimination trials. During the discrimination test, participants were asked to listen to the two utterances, and to decide whether they were the "same word", that is, having the same meaning, in the language. Responses were given on a paper answer sheet (Appendix B).

5.2.3.2 Musical Ear Test

In order to assess the musicality of the groups, the MET was administered to all participants at both time points. Due to time constraints, the original published version of the MET (Wallentin et al., 2010), which assesses only the discrimination of *contour* and *interval*, but not *key*, was used.

5.2.4 Results

5.2.4.1 Musicality

In order to assess the pre-existing differences and changes in musicality between the groups, the MET was first examined separately. A' was calculated as described in Equation 3.1 for each violation type (*contour* and *interval*) for each subject. Scores were regressed on group (aural skills musician, amateur musician, nonmusician), session, and violation, with session and subject as random effects; the results of this model are summarized in Table 5.3. MET scores are described in Table 5.4 and summarized graphically in Figures 5.1, 5.2, and 5.3.

$$A' = 1/2 + \frac{(hit - fa) * (1 + hit - fa)}{4hit(1 - fa)} \quad (3.1)$$

	β	Std. Error	t	p
(Intercept)	0.94042	0.03259	28.855	
group[amateur]	-0.03831	0.05927	-0.646	
group[nonmusician]	-0.22929	0.06784	-3.380	0.0019
session	-0.00622	0.01960	-0.317	
viol[interval]	-0.09270	0.04374	-2.119	0.025
group[amateur]:session	0.01368	0.03564	0.384	
group[nonmusician]:session	0.07057	0.04080	1.730	0.0514
group[amateur]:viol[interval]	-0.11060	0.07955	-1.390	
group[nonmusician]:viol[interval]	-0.21056	0.09106	-2.312	0.0172
session:viol[T.interval]	-0.03624	0.02766	-1.310	
group[amateur]:session:viol[interval]	0.04356	0.05031	0.866	
group[nonmusician]:session:viol[interval]	0.05062	0.05759	0.879	

Table 5.3: Estimates of fixed effects on MET scores in Experiment 4.

	pretest	posttest	average
aural skills musician	0.888 (0.048)	0.878 (0.067)	0.883 (0.058)
amateur musician	0.849 (0.070)	0.885 (0.052)	0.867 (0.061)
nonmusician	0.719 (0.142)	0.805 (0.090)	0.762 (0.116)

Table 5.4: Mean A' scores (SD in parentheses) for the MET in Experiment 4.

Aural skills musicians scored higher on the MET overall than nonmusicians ($\beta = -0.22929$, $p < 0.005$), and this difference was greatest on *interval* violations ($\beta = -0.21056$, $p < 0.05$). Aural skills musicians and amateur musicians did not differ significantly. Post-hoc tests of the aggregate MET scores (averaged over violation type and session) indicate that nonmusicians score significantly worse than aural-skills ($p_{adj} < .001$) and amateur musicians($p_{adj} < .001$). Changes in MET performance between the pretest and posttest did not appear to be significant for any group.

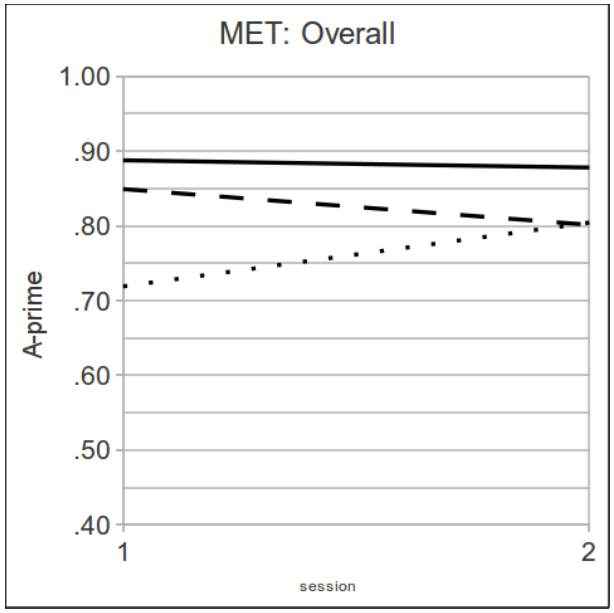

Figure 5.1: Overall MET performance by group (solid=aural skills musicians, dashed=amateur musicians, dotted=nonmusicians) in Experiment 4.

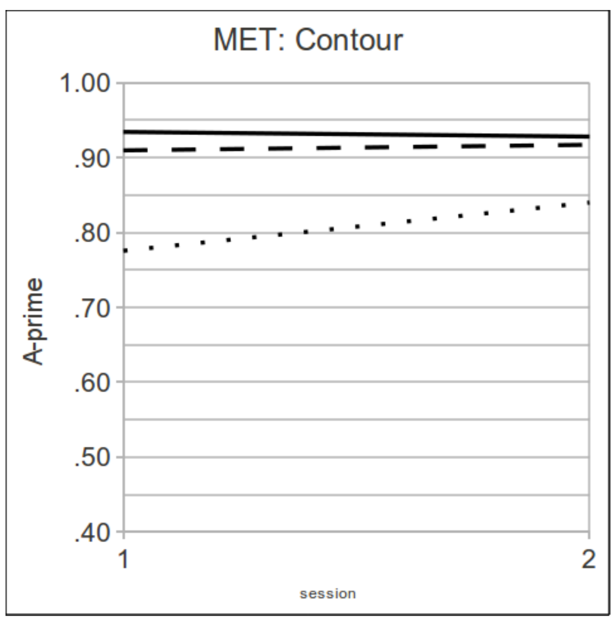

Figure 5.2: MET *contour* sensitivity by group (solid=aural skills musicians, dashed=amateur musicians, dotted=nonmusicians) in Experiment 4.

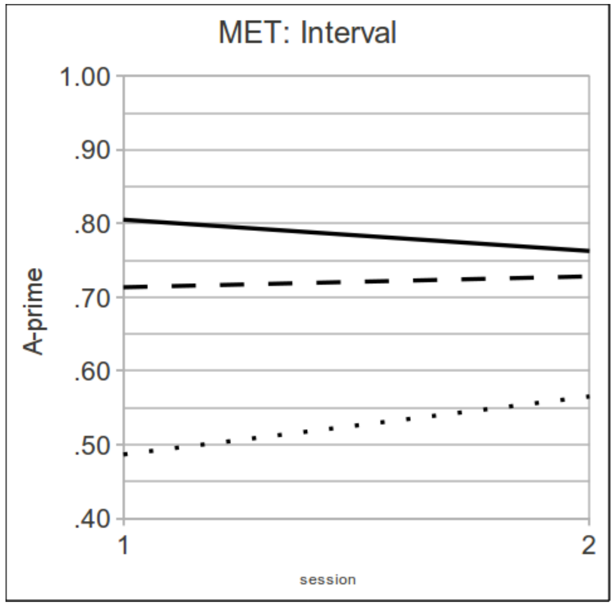

Figure 5.3: MET *interval* sensitivity by group (solid=aural skills musicians, dashed=amateur musicians, dotted=nonmusicians) in Experiment 4.

5.2.4.2 Tone Discrimination

The results of the tone discrimination test were analyzed similarly to the results of the MET; that is, as a signal detection task in which the signal participants must detect is the difference in tone category between two words. A' was calculated as described in Equation 3.1 for each contrast for each subject. The false alarm rate used for calculating each A' includes as "same" trials both tones of the contrast (*e.g.*, Tone 1 false alarm rate and Tone 2 false alarm rate for the Tone 1–Tone 2 contrast).

The results of the tone discrimination pretest are summarized in Figure 5.4 and Table 5.5. Participants performed well on almost all contrasts ($A' > .9$), except for the Tone 2–Tone 3 contrast, which is known to be difficult for learners of Mandarin (Shen & Lin, 1991; Wang et al., 1999; Wong, Skoe, et al., 2007). This is likely due to the low variability of the stimuli in the test; because the words were recorded from the same speaker in the same context, listeners did not need to normalize to the speaker or possess robust tonal categories. Discrimination could be done based on relatively superficial phonetic differences, rather than categorical encoding.

Post-hoc t-tests of the pretest confirmed that the Tone 2–Tone 3 contrast was discriminated significantly more poorly than the other contrasts ($p_{adj} < .001$) for all groups, and that the other contrasts did not differ significantly from one another. Due to the ceiling effects on other contrasts, subsequent analyses will focus only on the Tone 2–Tone 3 contrast.

contrast	1x2	1x3	1x4
aural skills musician	0.986 (0.025)	0.985 (0.049)	0.953 (0.052)
amateur musician	0.966 (0.072)	0.982 (0.031)	0.934 (0.077)
nonmusician	0.940 (0.114)	0.973 (0.044)	0.842 (0.146)
contrast	2x3	2x4	3x4
aural skills musician	0.820 (0.187)	0.976 (0.066)	0.987 (0.034)
amateur musician	0.642 (0.312)	0.957 (0.081)	0.976 (0.052)
nonmusician	0.472 (0.260)	0.935 (0.079)	0.968 (0.043)

Table 5.5: Mean A' scores (SD in parentheses) for pretest discrimination of Tone 2–Tone 3 in Experiment 4.

Discrimination scores for Tone 2–Tone 3 were regressed on group (aural skills musician, amateur musician, nonmusician) and session, with session and subject as random effects. The results of this model are summarized in Table 5.6. Tone discrimination pretest and posttest scores for Tone 2–Tone 3 are described in Table 5.7 and Figure 5.5.

Aural skills musicians performed better overall than nonmusicians ($\beta = -0.415575$, $p < 0.005$) and amateur musicians ($\beta = -0.241547$, $p < 0.05$). Figure 5.5 shows an increasing trend in A' from pretest to posttest for all groups, but only nonmusicians improved their discrimination of the Tone 2–Tone 3 contrast significantly from pretest to posttest ($\beta = 0.171129$, $p_{adj} = 0.01$).

	β	Std. Error	t	p
(Intercept)	0.913931	0.056315	16.229	
group[amateur]	-0.241547	0.102420	-2.358	0.02
group[nonmusician]	-0.415575	0.117229	-3.545	0.0027
session	-0.002649	0.029728	-0.089	
group[amateur]:session	0.067445	0.054067	1.247	
group[nonmusician]:session	0.171129	0.061884	2.765	0.01

Table 5.6: Estimates of fixed effects on discrimination of Tone 2–Tone 3 in Experiment 4.

	prestest	posttest	Δ p
aural skills musician	0.820 (0.187)	0.855 (0.203)	
amateur musician	0.642 (0.312)	0.728 (0.202)	
nonmusician	0.472 (0.260)	0.650 (0.324)	0.01

Table 5.7: Mean A' scores (SD in parentheses) for the discrimination of Tone 2–Tone 3 in Experiment 4.

5.2.4.3 Discrimination and Bias

Although discrimination scores (A') appear relatively flat for the musician groups, other trends between pretest and posttest emerge when the subcomponents of this score are examined. Specifically, there were changes in both accuracy/hit rate (correct discrimination of "different" trials) and false alarm rate (incorrect discrimination of "same" trials). Hit and false alarm rates are summarized in Tables 5.8 and 5.9 and Figures 5.6 and 5.7, respectively.

Despite the flatter discrimination scores for the musician groups, as compared to the nonmusicians, hit rates did increase for the amateur musician and aural skills

	pretest	posttest
aural skills musician	77.5% (17.2)	82.5% (19.6)
amateur musician	64.1% (27.3)	69.9% (21.4)
nonmusician	48.1% (19.9)	63.9% (27.3)

Table 5.8: Mean hit rates (SD in parentheses) for the Tone 2–Tone 3 contrast in Experiment 4.

	pretest	posttest
aural skills musician	7.8% (17.3)	9.8% (19.6)
amateur musician	3.4% (27.3)	5.6% (21.4)
nonmusician	9.0% (19.9)	6.2% (27.3)

Table 5.9: Mean false alarm rates (SD in parentheses) for the Tone 2–Tone 3 contrast in Experiment 4.

	pretest	posttest
aural skills musician	0.5498 (0.346)	0.3218 (0.479)
amateur musician	0.8811 (0.439)	0.7599 (0.480)
nonmusician	0.8327 (0.000)	0.7915 (0.333)

Table 5.10: Mean response bias scores (B''_D; SD in parentheses) in Experiment 4.

musicians (Figure 5.6); however, false alarm rates increased for these groups as well, while that for the nonmusicians fell (Figure 5.7), counteracting the effect of the hit rates on the discrimination scores of the musician groups.

These trends were analyzed through a measure of response bias ($B''D$; Equation 5.1). B''_D ranges from -1 to 1, with -1 indicating a strong liberal bias (a tendency to answer "different"), 1 indicating a strong conservative bias (a tendency to answer "same"), and 0 indicating no response bias. Bias scores for the three groups are summarized in Table 5.10 and Figure 5.8.

$$B''_D = \frac{(1 - hit) * (1 - fa) - hit * fa}{(1 - hit) * (1 - fa) + hit * fa} \tag{5.1}$$

Bias scores for Tone 2–Tone 3 were regressed on group (aural skills musician, amateur musician, nonmusician) and session, with session and subject as random effects. The results of this model are summarized in Table 5.11. Results indicate that aural skills musicians, but not the other groups, became less conservative in their response tendency from pretest to posttest ($\beta = -0.33218$, $p < 0.01$). While all groups displayed a conservative bias overall (a tendency to classify tone pairs as "same"), the aural skills group responded less conservatively (classified more pairs as "different") on the posttest, which was driven both by greater more discrimination of "different" trials and by a higher false alarm rate.

	β	Std. Error	t	p
(Intercept)	0.66743	0.18562	3.596	
group[amateur]	-0.47394	0.33759	-1.404	
group[nonmusician]	-0.06061	0.38641	-0.157	
session	-0.33218	0.11579	-2.869	0.0083
group[amateur]:session	0.21561	0.21058	1.024	
group[nonmusician]:session	0.18537	0.24103	0.769	

Table 5.11: Estimates of fixed effects on response bias in discrimination of Tone 2–Tone 3 in Experiment 4.

5.2.5 Discussion

5.2.5.1 Pre-existing Group Differences

5.2.5.1.1 Differences in Musicality

Differences between the aural skills musicians, amateur musicians, and nonmusicians at pretest are consistent with known effects of musical experience (Wallentin et al., 2010). The two groups with the greatest degree of musical experience (aural skills musicians and amateur musicians) recorded higher indicators of musicality as measured by the MET. This is unsurprising, given that the purpose of the MET is to distinguish between individuals with higher and lower levels of musical achievement or aptitude (Wallentin et al., 2010). There was no significant difference between the two musician groups, suggesting that the amateur musicians are a suitable control group for the musicians undergoing aural skills training.

5.2.5.1.2 Differences in Tone Perception

The three groups did not begin at the same level of performance on the tone discrimination test. Aural skills musicians outperformed nonmusicians, which is consistent with demonstrated effects of musicianship on tone perception (Kraus &

Chandrasekaran, 2010, *inter alia*). The aural skills musicians also outperformed the amateur musician group at pretest, albeit by a smaller margin. This indicates that there may be a difference in musicality between the musician groups which were not revealed by the MET. This difference is relatively small, however, and the two musician groups are more similar to each other than they are to the nonmusician groups; therefore, analysis of the effects of aural skills training, in comparison to the amateur musician control group, will proceed with caution. Future studies using random assignment to training or control paradigms will attempt to address this shortcoming.

5.2.5.2 Effects of Aural Skills Training

5.2.5.2.1 Effects on Musicality

Only the nonmusician group demonstrated improvement in MET performance from pretest to posttest. Because this group engaged in no structured musical activity in the intermediate period, this improvement is assumed to represent a retest effect. That the two musician groups, and specifically the group undergoing aural skills training, did not exhibit such improvement is attributed to the fact that the musician groups scored near the ceiling of the MET at pretest, and thus the test was not sensitive to any effect in these groups, whether due to retest or training. Because a purpose of the MET is to distinguish between musicians and nonmusicians, a more difficult version of the test must be developed to distinguish between musicians with more and less advanced perceptual abilities.

5.2.5.2.2 Effects on Tone Perception

For the contrast considered (Tone 2–Tone 3), a change in sensitivity (as measured by A') was again more evident for the nonmusician group, which can be partially attributed to a retest effect within the group scoring lowest at pretest (the nonmusicians), combined with a ceiling effect on those scoring highest at pretest (the aural skills and amateur musicians).

As with the MET, the tone discrimination test used in Experiment 4 was not an ideal instrument for measuring change in tone perception in the groups studied. This is because the stimulus set employed has a low level of variability due to the fact that it is composed of tokens recorded from only a single speaker. This makes the test much easier, because such stimuli can be distinguished on the basis of raw acoustic differences (rather than tone category membership) more than can tokens from a high-variability stimulis set including multiple speakers. This difference between acoustic and categorical discrimination gives a clue toward the interpretation of other differences between the groups. When the data are examined in more detail, the picture becomes more complicated, but an effect emerges which is unique to the aural skills group.

The purpose of using A' as the measure of discrimination is to take all trials completed by a participant involving the tones of interest, rather than accuracy (or hit rate), which only includes performance on "different" trials. Hit rate rose for all groups. It rose most of all for the nonmusicians, who began below chance (Table 5.8). The other component of A', false alarm rate, fell for nonmusicians (Table 5.9), resulting in a large improvement in discrimination score. The changes in hit and false alarm rates for this group are assumed to be retest effects, because this group engaged in no musical or language learning activity in the period between pretest and posttest.

For the other two groups, the false alarm rate rose, resulting in a flattening of A' scores for these groups, despite an increase in hit rate. Using the overall measure of bias (B''_D), the aural skills musician group was the only group for whom response bias changed significantly between pretest and posttest; musicians undergoing aural skills training became less conservative in their responses—they more often labeled tone pairs as “different” in the posttest, resulting in a greater proportion of both false alarms and hits.

This appears to suggest a worsening of tone performance by the aural skills group, the group for which improvement in tone discrimination was hypothesized. However, it is important to consider the context of the training they received, and the framework in which phonological learning is hypothesized to take place. Participants in Experiment 4 did not actually learn anything about the tonal categories of Mandarin; they were not told anything about the tones themselves and received no feedback on the test. Further, the “same” trials in the tone discrimination test contained *different* tokens of the same pseudoword, so they were not acoustically identical. While A' indicates sensitivity to a categorical (phonological) distinction, a shift in response bias (B''_D) indicates a change in the ability to discriminate speech tokens differing in pitch content on the basis of acoustic or phonetic differences.

This is consistent with the phonetic–phonological–lexical continuum of language learning (Wong & Perrachione, 2007); the effect seen here indicates changes at the phonetic level for the aural skills musicians, which is the level which the hypothesized mapping between features of lexical tones and melodies (Table **??**) is based upon. In order to become more like native speakers, learners of a language must simultaneously increase their hits and lower their false alarms in the perception of phones. This requires reinforcement of arbitrary categorical boundaries, and the musical experiences engaged in by the aural skills group should not be expected to transfer to this “phonological” level of language perception. This would be part of the domain-specific “representation” network, rather than the domain-general “resource” network referred to by Patel (2008a). Within the general framework of RHT, transfer between linguistic and musical experience occurs at this lower level, and differences in phonetic perception emerging from musical experience may underlie the advantages in “higher” phonological and lexical learning by musicians

(Alexander et al., 2005; Wong & Perrachione, 2007; Chandrasekaran et al., 2010, *inter alia*).

These results do not match expected findings as articulated in Hypothesis 4′, but suggest some perceptual changes resulting from aural skills training. Because the content of the aural skills training was not controlled in this study, these findings do not directly address the mapping between melodic and tonal properties described in Table **??**. However, aural skills training is known to involve the discrimination and identification of melodic *contour* and *interval*, and the tonal contrast examined (Tone 2–Tone 3) is primarily based on *slope*. Although Tone 2 is typically described as 'rising', both Tone 2 and Tone 3 have a concave shape. The degree of *slope* in the rising and falling components and the timing between them is important to this contrast (Shen & Lin, 1991). Tone 3 also includes a creaky voice quality, but all voice quality correlates were removed in the stimuli used in Experiment 4. These findings are at least indirectly supportive of this mapping, in that learning to perceive musical intervals leads to changes in the way a lexical tone which shares acoustic properties are perceived.

Furthermore, the high level of musicality (near ceiling) of both the aural skills musicians and the amateur musicians did not allow for a direct comparison of explicit perceptual training to music participation. The application of a similar learning paradigm to participants from across a musicality spectrum, especialy to nonmusicians, woud more directly address the predictions of the OPERA hypothesis.

5.3 Summary

Musicians undergoing ear-training and amateur musicians with a similar level of musicality who participated in musical activities outperformed nonmusicians on tests of melody and lexical tone discrimination administered at two time points. Nonmusicians improved on both tests between time points (although they still did not reach the level of the musicians), which is attributed to a retest effect. Neither musician group improved on either test, which is likely due to a ceiling effect.

A change in response bias was observed only among musicians undergoing aural skills training; those who underwent training were more likely to discriminate between (always non-identical) tokens in the tone discrimination posttest. This is argued to indicate an improvement in phonetic sensitivity without concomitant phonological learning.

This effect was not observed in the amateur musician group, who did not undergo any special training between pretest and posttest. Given the similar baseline level of musicality between the aural skills musicians and amateur musicians, compared to the nonmusician group, the kind of aural skills training examined here appears to have had an effect on lexical tone perception beyond simply participating in music, or a general level of musicianship, and this effect appears to occur specifically at the phonetic level.

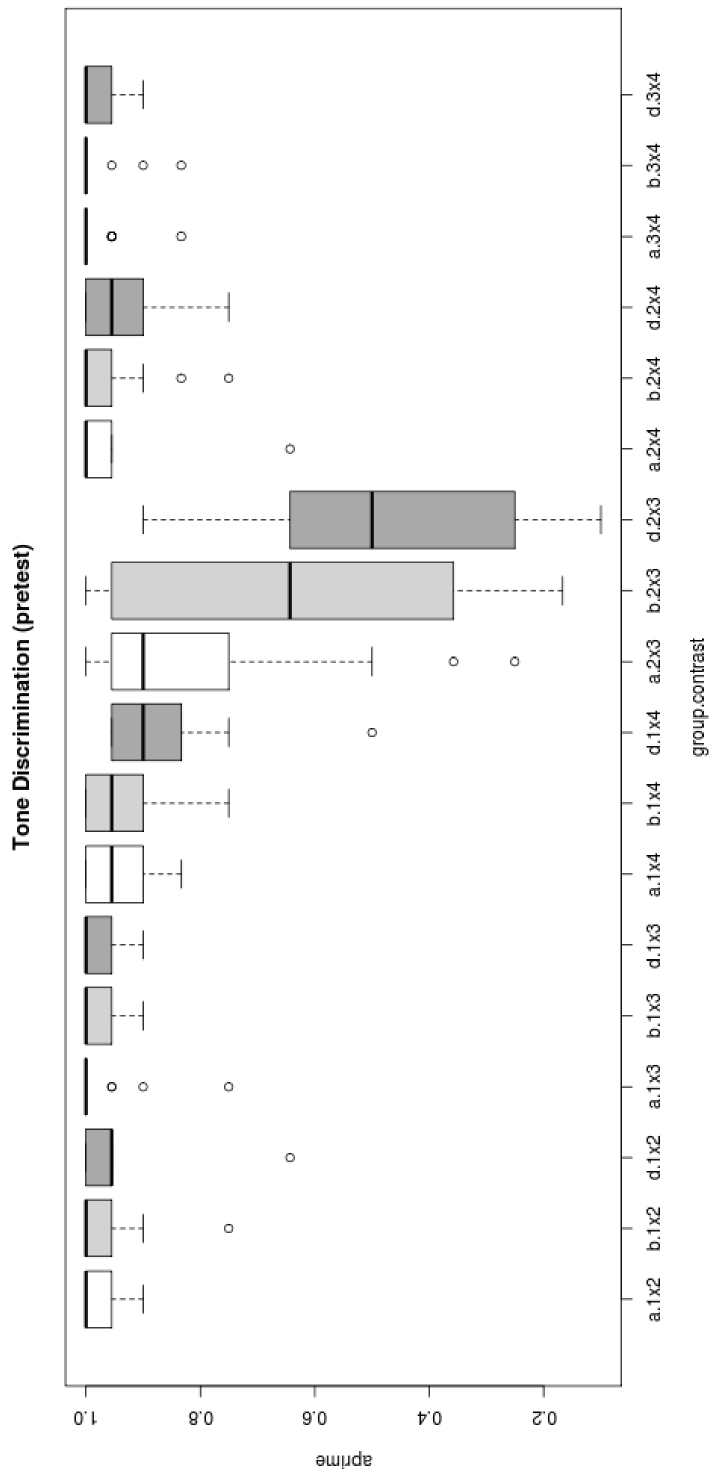

Figure 5.4: Pretest Mandarin tone discrimination by contrast and group in Experiment 4.

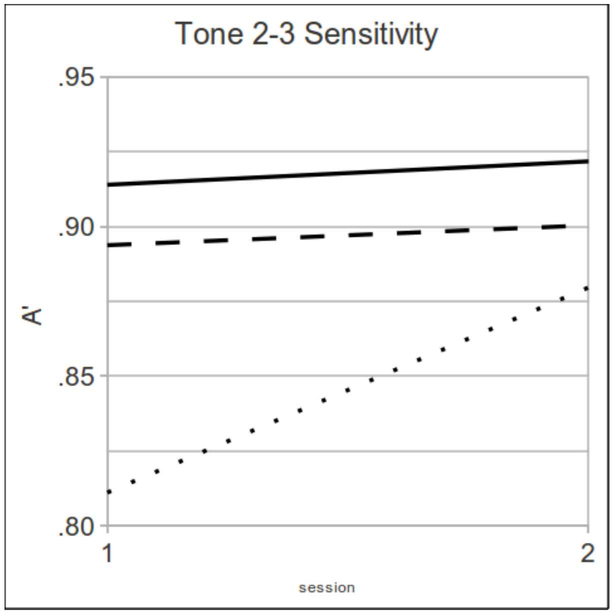

Figure 5.5: Changes in sensitivity (A') to the Tone 2–Tone 3 contrast by group (solid=aural skills musicians, dashed=amateur musicians, dotted=nonmusicians) in Experiment 4.

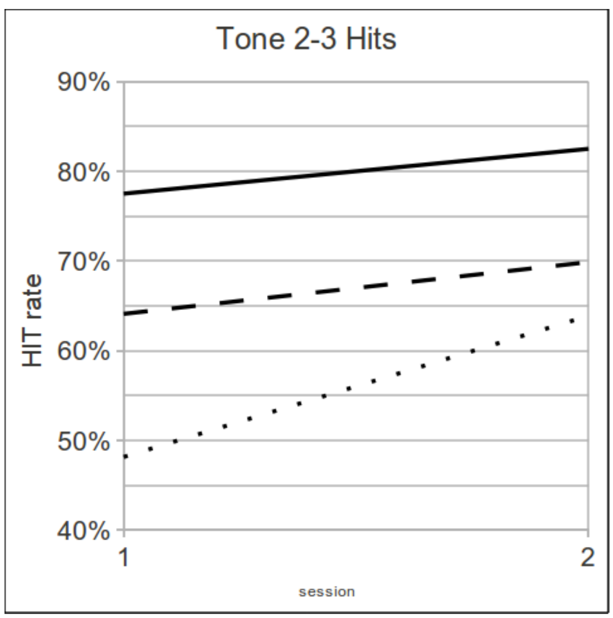

Figure 5.6: Changes in hit rate (correct discrimination of the Tone 2–Tone 3 contrast) by group (solid=aural skills musicians, dashed=amateur musicians, dotted=nonmusicians) in Experiment 4.

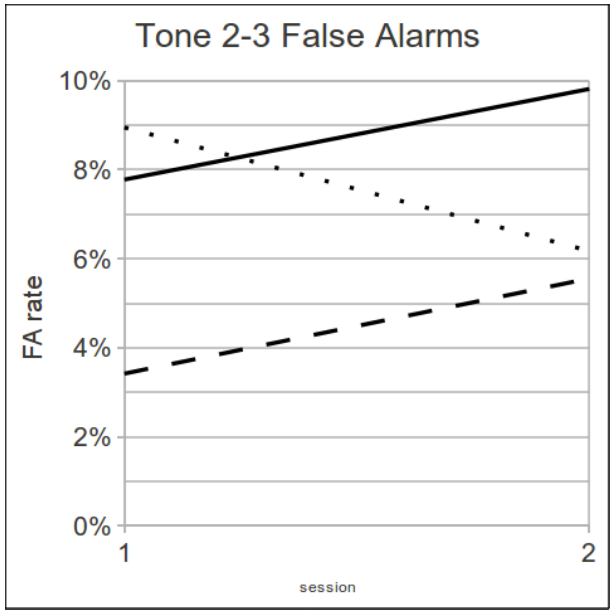

Figure 5.7: Changes in false alarm rate for Tone 2 and Tone 3 by group (solid=aural skills musicians, dashed=amateur musicians, dotted=nonmusicians) in Experiment 4.

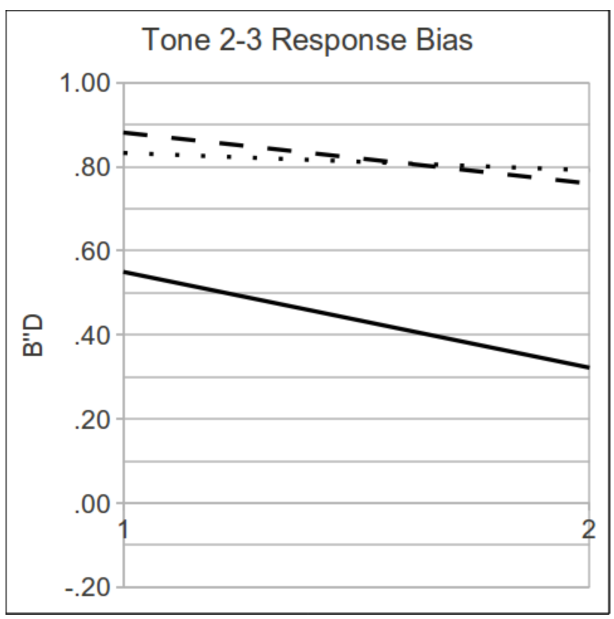

Figure 5.8: Changes in response bias for Tone 2 and Tone 3 by group (solid=aural skills musicians, dashed=amateur musicians, dotted=nonmusicians) in Experiment 4.

Chapter 6

GENERAL DISCUSSION AND CONCLUSIONS

Chapters 3–5 described a series of experiments designed to test specific hypotheses following from the General Hypothesis and the mapping between properties of tone and melody (Table **??**) developed in Chapter 2. This chapter will summarize and integrate those results, discuss their implications for this framework, and identify outstanding questions.

6.1 Summary of Experiments

Experiment 1 examined the effects of native language on the discrimination of musical melody. A melodic discrimination test was administered to native speakers of Mandarin, Yoruba, and English. Mandarin speakers exhibited greater sensitivity to changes in melodic *contour* and *interval* than did English speakers. Yoruba speakers exhibited greater sensitivity to changes in *interval* than English speakers. The three groups did not differ in sensitivity to *key*. Ther performance of the Mandarin group matched hypotheses, but the Yoruba group did not; Yoruba speakers were expected to outperform English speakers only on *contour*, but they differed from English speakers only on sensitivity to *interval*.

Experiment 2 and Experiment 3 attempted to examine the effects of lexical tones in a second language learned in adulthood by examining native English-speaking learners of Mandarin in longitudinal and cross-sectional administrations of a melodic discrimination test. These learners did not demonstrate improvement in perception of melody, contrary to expectation. This may be partially attributable to the relatively low levels of Mandarin proficiency achieved by these participants, or to differences between comprehensive language study and previously reported effects of laboratory-based phonetic training, but the possibility remains that second language tone experience does not affect music perception in the same manner or to the same degree as native language experience.

Experiment 4 examined the effects of different kinds of musical experience on the perception of lexical tones by English speakers. English-speaking music students completed a Mandarin tone discrimination test before and after undergoing aural skills training. Their performance was compared to that of amateur musicians and nonmusicians who did not undergo such training. Although some improvement in discrimination of the Tone 2–Tone 3 contrast was evident in both the amateur and nonmusician groups, this is most likely attributable to a retest effect, as only the nonmusician group improved significantly. The aural skills group did not improve their tone perception compared to the other groups, likely due to the fact that they scored near ceiling, even at pretest.

However, there were changes in tone performance found only among the aural skills group. The aural skills musicians also showed an increase in response bias after training—a tendency to discriminate within-category tone tokens as "different". This change in response bias by the aural skills group is not an "improvement" in tone perception *per se*, but it is consistent with a change in pitch percpeption at the phonetic or acoustic level, the level at which the tone–melody mapping (Table **??**) is formulated.

Although this seems counterproductive from the standpoint of language learning, it is consistent with the hypothesized effects of music-induced sensory tuning. Because these participants were not learning or receiving any feedback about the tonal categories of Mandarin, a change in phonetic-level sensitivity to pitch *direction* or *slope* should not lead to categorical perception on its own, but may prime the system for more efficient learning of such categories later by pushing musicians further along the phonetic–phonological–lexical continuum. A change in response bias may thus be a more sensitive indicator of early changes in tone perception than sensitivity (A') or accuracy.

6.2 First and Second Language Effects

It appears that the Mandarin learners in Experiment 2 and Experiment 3 did not come to resemble the native Mandarin speakers in Experiment 1 in terms of their melody perception performance, as expected.

Neural changes at the level of the auditory brainstem have been demonstrated in response to second language tone learning (Song et al., 2008). This tuning occured after Mandarin lexical training of dozens of hours over a few weeks, and resembled neural differences between Mandarin and English speakers (Chandrasekaran, Krishnan, & Gandour, 2009) and between musicians and nonmusicians (Wong, Skoe, et al., 2007). Participants who underwent training displayed more robust representation of pitch movement, and it was these effects which led to the expectation that students studying Mandarin would exhibit the reverse effect on melody perception.

The neural effects of language learning were not examined in Experiment 2 or Experiment 3, so an explanation for this difference between first and second language tone experience must remain speculative at this stage, but there appear to be two possible reasons for this null result.

First, that no neural tuning in response to tone learning occured among the participants, meaning that a prerequisite for improved melodic perception was not obtained. This is certainly possible, given the relatively low proficiency of the speakers in these experiments, and the differences between comprehensive language learning and laboratory based training. However, given the results of Song et al.

(2008), it is unlikely that highly profient (fluent) second language speakers of Mandarin would not display neural differences compared to monolingual English speakers. This could be verified by neural studies of similar groups of learners, and by the inclusion of more advanced learners in Experiment 2 and Experiment 3.

A second possible cause of the unexpected lack of effects in the second language studies is that, although participants learned tones and underwent neural changes similar to previous studies, the lack of change in melody performance is due to a failure to integrate these neural changes with the musical task. That is, changes in acoustic/phonetic perception are not sufficient to cause the phonological/lexical-type effects measured by the MET. This could be due to proficiency or duration of language study, as noted above, or some other difference between the experience of native and second language tone learners. This should be tested through the inclusion of a wider array of participants, as well as intermediate versions of the MET focusing on identification or discrimination of isolated contours and intervals. These are less complex, and may be more likely to be affected at an early stage of learning than the complex melodies tested by the MET.

6.3 Musical and Linguistic Effects

The Mandarin tone contrast affected in Experiment 4 (Tone 2–Tone 3) is based largely on *slope*; each tone has a concave shape, and the degree of *slope* in the rising and falling components and the timing between them are important for distinguishing them (Shen & Lin, 1991). Tone 3 also includes a creaky voice quality, but all voice quality correlates were removed in the stimuli used in Experiment 4.

In Experiment 1, Mandarin and English speakers differed significantly on the perception of melodic *interval*, the melodic dimension argued to correspond to the tonal property *slope*. This suggests but does not conclusively establish a parallel between the effects of linguistic experience seen in Experiment 1 and those of musical ear training seen in Experiment 4.

Interval and *contour* were not manipulated separately in Experiment 4; the aural skills musicians in the study followed a standard currculum, which included study of melodic structures like *contour* and *interval*, as well as other music theoretic concepts. A laboratory-based aural skills training paradigm, allowing more direct manipulation of these properties, should be conducted to establish this effect more conclusively.

It is also unknown whether the effects seen in Experiment 4 also apply to the other tone contrasts of Mandarin, which were not examined due to ceiling effects. According to the General Hypothesis, perceptual tuning to musical properties such as *interval* should lead to effects uniquely on *slope*, and thus, Mandarin tone contrasts not as reliant on this property as Tone 2–Tone 3 (*e.g.*, Tone 1–Tone 2) should not be affected to the same degree, nor should tonal constrasts form languages whose tones are not reliant on these properties.

As discussed in Chapter5, refinement of the tone discrimination test through increased variability in the stimulus set will allow better examination of the other tone contrasts of Mandarin. In addition, the incorporation of tones from other languages could provide stronger evidence and help to generalize the links between melodic and tonal properties.

6.4 Hypotheses Revisited

The General Hypothesis proposed in Chapter 2, based on principles derived from RHT (Ahissar et al., 2009) and SSCLMH (Patel, 2008a) was generally supported by the key findings of Experiment 1 and Experiment 4, in that

1. tone language experience leads to improvement in some, but not all components of melody perception; and

2. aural skills training leads to narrow changes in the perception of lexical tones.

The General Hypothesis was not extended to second language tone experience on the basis of Experiment 2 and Experiment 3, and the results Experiment 4 were unclear, further research is needed in this area.

While the general idea of a correspondance between stuctural properties of tone and melody specific hypotheses is generally consistent with the findings presented, the specific mappings proposed (Table **??**) were supported to varying degrees.

tone	**melody**
height	*key*
direction	*contour*
slope	*interval*

As discussed in Section 6.3, the results of Experiment 1 and Experiment 4 together suggest that *slope* and *interval* form a relatively good match. The mapping between *direction* and *contour* was supported by the fact that Mandarin speakers outperformed English speakers on the *contour* component of the MET. The performance on the Yoruba group in Experiment 1 complicates the picture slightly. Yoruba speakers did not differ from English speakers on *contour*, as predicted. This could be due to the limitations MET, or to improper assumptions about the facts of Yoruba. Both of these issues should be addressed before the mapping between *direction* and *contour* is revised.

Although not directly incorporated into the General Hypothesis, which states that music–language crossover effects should be bidirectional, the principles of the OPERA hypothesis predict that such effects are moderated by the context of learning, and that certain characteristics of musical training could result in assymmetries in the magnitude of such effects in each direction. The constuction of similar tone and melody learning paradigms would allow the systematic examination of these characteristics, and would help to further disentagle those effects which are due to shared properties of language and music, and which are due to the context of learning.

Although *height* and *key* were linked in the mapping, this particular correspondance has not been tested directly, in that no difference was predicted for any of the examined groups based on these dimensions. The question remains as to whether the differences seen between speakers of various languages in their reliance on *height* in tone perception (Sections 2.1.1.3, 2.1.1.5.2) have a corresponding effect on music perception.

If there is a fundamental distinction between the static and dynamic properties included in the mapping, and static properties do reflect a more general auditory ability than dynamic properties do (Gandour & Harshman, 1978), then this particular mapping may be less affected by experience, at least at the phonetic level. This is suggested by the high performance of all groups in Experiment 4 on *key* discrimination, and the commonality of absolute pitch memory abilities in the general population (Levitin, 1994). The inclusion of speakers of additional languages (*e.g.*, Cantonese) may help to answer this question.

6.5 Extending the Mapping

Some of the inconsitencies between the initially proposed mappings and the experimental results may be alleviated by the addition of new tonal and melodic properties. One clue to how the mapping may be extended comes from evidence demonstrating the role of curvilinear pitch changes in tone perception (Chandrasekaran, Krishnan, & Gandour, 2007a; Y. Xu, Krishnan, & Gandour, 2006; Krishnan et al., 2009). This curvilinear property is a higher-order dimension of tone, which is not independent from *slope*. It is also not clear what musical elements, if any, would map to this tonal property.

In addition to adding and revising mappings between properties of F_0, the mapping should be extended to include other kinds of pitch information, such as harmonics (Krishnan et al., 2005, 2009; Liang, 1963; Fu et al., 1998; Abramson, 1973; Liu & Samuel, 2004; Rasch & Plomp, 1982; Burns & Ward, 1982; Tramo et al., 2003), and other spectral and temporal acoustic properties which may be shared by language and music (Kong & Zeng, 2006; L. Xu & Pfingst, 2003; Krishnan, Gandour, & Bidelman, 2010a).

6.6 Summary and Conclusions

The ideas presented in this dissertation represent an attempt to clarify the cognitive relationship between language and music by focusing on a salient case of overlap between the two complex domains: the use of pitch in lexical tone and musical melodies. A General Hypothesis was developed on the basis of a theory of perceptual learning (RHT), and specific predictions were made on the basis of the typology of tone languages and facts about native and second language perception and the development of musicianship.

Although these predictions were only partially supported by the experimental findings, the results established a framework which furthers our understanding of the relationship between language and music: music and language are built from similar acoustic information, and the architecture of the human auditory system and its mechanisms for plasticity allow for the transfer of learning between linguistic and musical tasks when their acoustic structures are similar. This project has also developed a suite of experimental tools which, along with this framework of hypotheses, will facilitate investigations which have the potential to unite and explain some cases of transfer between language and music, to make new predictions about the effects of language and music on the perceptual system, and to provide new perspective on parallels between the structure of music and language.

References

Abramson, A. S. (1973). Tonal experiments with whispered Thai. In A. Valdman (Ed.), *Papers in linguistics and phonetics to the memory of Pierre Delattre* (p. 31-44). The Hague: Mouton.

Abramson, A. S. (1975). The tones of Central Thai: Some perceptual experiments. In J. G. Harris & J. R. Chamberlain (Eds.), *Studies in Tai linguistics in honor of William J. Gedney.* Bangkok: Central Institute of English Language, Office of State Universities.

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

Abramson, A. S. (1979). The noncategorical perception of tone categories in Thai. In B. Lindblom & S. Öhman (Eds.), *Frontiers of speech communication research* (pp. 127–134). London: Academic Press.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464.

Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1515), 285–299.

Akinlabi, A. (2001). Yoruba. In J. Garry & C. Rubino (Eds.), *Facts about the world's languages: An encyclopedia of the world's major languages, past and present* (pp. 836–841). New York: The H.W. Wilson Company.

Akinlabi, A., & Liberman, M. (2001). Tonal complexes and tonal alignment. In M. Kim & U. Strauss (Eds.), *Proceedings of the North East Linguistic Society* (Vol. 31, pp. 1–20). Amherst, MA: GLSA.

Alexander, J. A., Bradlow, A. R., Ashley, R. D., & Wong, P. C. (2011). Music-melody perception in tone-language and non-tone-language speakers. In *Psycholinguistic representation of tone.* Hong Kong.

Alexander, J. A., Wong, P. C., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. In *Proceedings of the 9th European Conference on Speech Communication and Technology.* Lisbon.

Audacity Development Team. (2010). Audacity [Computer software manual].

Avelino, H. (2003). Categorical perception of phonemic tone in Yalálag Zapotec. In *Proceedings of the 15th International Congress of the Phonetic Sciences.*

Bartlett, J. C., & Dowling, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(3), 501–515.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using s4 classes [Computer software manual]. (R package version 0.999999-0)

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

Bergeson, T. R., & Trehub, S. E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science*, *13*(1), 72–75.

Besson, M., Chobert, J., & Marie, C. (2011). Language and music in the musician brain. *Language and Linguistics Compass*, *5*(9), 617–634.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listeners native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775.

Bidelman, G. M., & Krishnan, A. (2009). Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *Journal of Neuroscience*, *29*(42), 13165–13171.

Bigand, E. (2003). More about the musical expertise of musically untrained listeners. *Annals of the New York Academy of Sciences*, *999*, 304–312.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? a review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*(1), 100–130.

Boersma, P., & Weenink, D. (2010). *Praat: Doing phonetics by computer.* (Version 5.1.25)

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Lacerda, F., & Attapaiboon, C. (1996). Facilitation or attenuation in the development of speech mode processing? tone perception over linguistic contexts. In *Sixth Australian International Conference on Speech Science and Technology* (p. 587-592).

Burns, E. M., & Ward, W. D. (1982). Intervals, scales, and tuning. In D. Deutsch (Ed.), *The psychology of music* (p. 241-269). New York: Oxford University Press.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General*, *113*(3), 394–412.

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*(3-4), 195–210.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007a). Experience-dependent neural plasticity is sensitive to shape of pitch contours. *Neuroreport*, *18*(18), 1963–1967.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007b). Mismatch negativity to pitch contours is influenced by language experience. *Brain research*, *1128*(1), 148–156.

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

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

Connell, B. A. (2000). The perception of lexical tone in Mambila. *Language and Speech*, *43*(2), 163–182.

Connell, B. A., Hogan, J., & Rozsypal, A. (1983). Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *Journal of Phonetics*, *11*(4), 337–351.

Connell, B. A., & Ladd, D. R. (1990). Aspects of pitch realisation in Yoruba. *Phonology*, *7*(1), 1–29.

Cooke, P. (1992). Report on pitch perception experiments carried out in Buganda and Bosoga (Uganda). *African Music*, *7*(2), 119–125.

Dediu, D. (2010). A Bayesian phylogenetic approach to estimating the stability of linguistic features and the genetic biasing of tone. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*.

Dediu, D., & Ladd, D. R. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(26), 10944–10949.

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

Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception*, *21*(3), 339–356.

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *The Journal of the Acoustical Society of America*, *119*(2), 719.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341-354.

Edworthy, J. (1985). Interval and contour in melody processing. *Music Perception*, *2*(3), 375–388.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, *23*(4), 527–552.

Francis, A. L., & Ciocca, V. (2003). Stimulus presentation order and the perception of lexical tones in Cantonese. *The Journal of the Acoustical Society of America*, *114*(3), 1611.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*(2), 268–294.

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

Fu, Q. J., Zeng, F.-G., Shannon, R. V., & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *The Journal of the Acoustical Society of America*, *104*(1), 505–510.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience*, *16*(6), 1010–1021.

Gandour, J. T. (1978). Perceived dimensions of 13 tones: a multidimensional scaling investigation. *Phonetica*, *35*(3), 169–179.

Gandour, J. T. (1979a). Perceptual dimensions of Cantonese tones: a multidimensional scaling reanalysis of Fok's tone confusion data. *Southeast Asian Linguistic Studies*, *4*, 415–429.

Gandour, J. T. (1979b). Perceptual dimensions of tone: Thai. *Southeast Asian Linguistic Studies*, *3*, 277–300.

Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*(2), 149–175.

Gandour, J. T. (1998). Pitch processing in the human brain is influenced by language experience. *Neuroreport*, *9*(9), 2115–2119.

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

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *Journal of Neuroscience*, *23*(27), 9240–9245.

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

Gussenhoven, C. (2004). Pitch in language II: Tone. In *The phonology of tone and intonation* (p. 26-48). New York: Cambridge University Press.

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

Honorof, D. N., & Whalen, D. H. (2005). Perception of pitch location within a speaker's F0 range. *The Journal of the Acoustical Society of America*, *117*(4), 2193.

House, D. (1989). Perceptual constraints and tonal features. *Lund University Working Papers*, *35*, 113–120.

Hove, M., Sutherland, M., & Krumhansl, C. (2010). Ethnicity effects in relative pitch. *Psychonomic Bulletin & Review*, *17*, 310-316.

Hsieh, I.-H., & Saberi, K. (2008). Language-selective interference with long-term memory for musical pitch. *Acta Acustica united with Acustica*, *94*(4), 588–593.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., & Siebert, C. (2001). A perceptual interference account of acquisition difficulties for non-native phonemes. *Speech, Hearing and Language: work in progress*, *13*, 106–118.

Jiang, X., Bradley, E. D., Rini, R. A., Zeffiro, T., & Vanmeter, J. (2006). Categorization training leads to sharpening tuning of shape-specific tuning in the lateral occipital cortex and learning of category-selective representations in the prefrontal cortex. *Journal of Vision*, *6*(6).

Jiang, X., Bradley, E. D., Rini, R. A., Zeffiro, T., Vanmeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron*, *53*(6), 891–903.

Johnson, K. (2005). Speaker normalization in speech perception. In D. Pisoni & R. Remez (Eds.), *The handbook of speech perception* (pp. 363–389). Blackwell Publishers.

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics*, *35*(1), 104–117.

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*(4), 646–653.

Köhler, . J. (1987). Categorical pitch perception. In *Proceedings of the 11th International Congress of Phonetic Sciences* (Vol. 5, pp. 149–152).

Kong, Y.-Y., & Zeng, F.-G. (2006). Temporal and spectral cues in Mandarin tone recognition. *The Journal of the Acoustical Society of America*, *120*(5), 2830.

Kraus, N., & Banai, K. (2007). Auditory-processing malleability: Focus on language and music. *Current Directions in Psychological Science*, *16*(2), 105–110.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605.

Kraus, N., Skoe, E., Parbery-Clark, A., & Ashley, R. (2009). Experience-induced malleability in neural encoding of pitch, timbre, and timing. *Annals of the New York Academy of Sciences*, *1169*, 543–557.

Krishnan, A., & Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain and Language*, *110*, 135–148.

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

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

Krishnan, A., Swaminathan, J., & Gandour, J. T. (2009). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *Journal of Cognitive Neuroscience*, *21*(6), 1092–1105.

Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, *25*(1), 161–168.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* New York: Oxford University Press.

Lee, C.-Y. (2009). Identifying isolated, multispeaker Mandarin tones from brief acoustic input: a perceptual and acoustic study. *The Journal of the Acoustical Society of America*, *125*(2), 1125–1137.

Lee, J., Perrachione, T. K., Dees, T. M., & Wong, P. C. (2007). Differential effects of stimulus variability and learners' pre-existing pitch perception ability in lexical tone learning by native English speakers. In *Proceedings of the 16th International Congress of Phonetic Sciences* (p. 1589-1592).

Lee, K. M., Skoe, E., Kraus, N., & Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *Journal of Neuroscience*, *29*(18), 5832–5840.

Levitin, D. J. (1994). Absolute memory for musical pitch: evidence from the production of learned melodies. *Perception & Psychophysics*, *56*(4), 414–423.

Levitin, D. J., & Rogers, S. E. (2005). Absolute pitch: perception, coding, and controversies. *Trends in Cognitive Sciences*, *9*(1), 26–33.

Liang, Z. A. (1963). The auditory perception of Mandarin tones. *Acta Physiologica Sinica*, *26*, 85–91.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368.

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and speech*, *47*(2), 109–38.

Lynch, M. P., & Eilers, R. E. (1991). Children's perception of native and nonnative musical scales. *Music Perception*, *9*(1), 121–131.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.*

Cambridge: Cambridge University Press.

Maddieson, I. (1978). Universals of tone. In J. H. Greenberg (Ed.), *Universals of human language* (Vol. 2, pp. 335–365). Stanford, CA: Stanford University Press.

Maddieson, I. (1991). Tone spacing. *York Papers in Linguistics*, *15*, 28.

Maddieson, I. (2005). Tone. In M. Haspelmath, M. S. Dryer, & D. Gil (Eds.), *The world atlas of language structures* (chap. 13). Oxford: Oxford University Press.

Massaro, D. W., Kallman, H. J., & Kelly, J. L. (1980). The role of tone height, melodic contour, and tone chroma in melody recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(1), 77–90.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, *106*(3), 1367–1381.

McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology*, *18*(4), 452–463.

McMullen, E., & Saffran, J. R. (2010). Music and language: A developmental comparison. *Music Perception*, *21*(3), 289–311.

Merrill, J., Sammler, D., Bangert, M., Goldhahn, D., Lohmann, G., Turner, R., et al. (2012). Perception of words and pitch patterns in song and speech. *Frontiers in Psychology*, *3*, 1-13.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*(1-2), 36–47.

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

Münte, T. F., Altenmüller, E. O., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, *3*(6), 473–478.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(40), 15894–15898.

Oechslin, M. S., Meyer, M., & Jäncke, L. (2010). Absolute pitch–functional evidence of speech-relevant auditory acuity. *Cerebral Cortex*, *20*(2), 447–455.

Pallier, C. (2002). *Computing discriminability and bias with the R software.*

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, *392*(6678), 811–814.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport*, *12*(1), 169–174.

Pantev, C., Ross, B., Fujioka, T., Trainor, L. J., Schulte, M., & Schulz, M. (2003). Music and learning-induced cortical plasticity. *Annals of the New York Academy of Sciences*, *999*, 438–450.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681.

Patel, A. D. (2008a). *Language, music, and the brain.* Oxford: Oxford University Press.

Patel, A. D. (2008b). Sound elements. In *Language, music, and the brain* (p. 7-94). New York: Oxford University Press.

Patel, A. D. (2009). Language, music, and the brain: a resource-sharing framework. In P. Rebuschat, M. Rohrmeier, J. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems.* Oxford: Oxford University Press.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? the OPERA hypothesis. *Frontiers in Psychology*, *2*(142).

Patel, A. D. (2012). The OPERA hypothesis: assumptions and clarifications. *Annals of the New York Academy of Sciences*, *1252*(1), 124–128.

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., & Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, 1–9.

Peng, S.-h. (1997). Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics*, *25*(3), 371–400.

Peretz, I., & Babaï, M. (1992). The role of contour and intervals in the recognition of melody parts: Evidence from cerebral asymmetries in musicians. *Neuropsychologia*, *30*(3), 277–292.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*, 58–75.

Pfordresher, P., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, & Psychophysics*, *71*, 1385-1398.

Pierrehumbert, J., & Steele, S. (1989). Categories of tonal alignment in English. *Phonetica*, *46*(4), 181–196.

R Core Team. (2012). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. (ISBN 3-900051-07-0)

Rammsayer, T., & Altenmüller, E. O. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception*, *24*(1), 37–48.

Rasch, R. A., & Plomp, R. (1982). The perception of musical tones. In D. Deutsch (Ed.), *The psychology of music* (p. 1-21). New York: Academic Press.

Schellenberg, E. G., & Trehub, S. E. (2008). Is there an Asian advantage for pitch memory? *Music Perception*, *25*(3), 241–252.

Schlaug, G., Jäncke, L., Huang, Y., Staiger, J., & Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia*, *33*(8), 1047–1055.

Schlaug, G., Jäncke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, *267*(5198), 699–701.

Schmithorst, V. J., & Wilke, M. (2002). Differences in white matter architecture between musicians and non-musicians: a diffusion tensor imaging study. *Neuroscience Letters*, *321*(1-2), 57–60.

Schneider, A. (2001). Sound, pitch, and scale: from "tone measurements" to sonological analysis in ethnomusicology. *Ethnomusicology*, *45*(3), 489.

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, *5*(7), 688–94.

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience*, *23*(13), 5545–5552.

Shahin, A., Roberts, L. E., & Trainor, L. J. (2004). Enhancement of auditory cortical development by musical experience in children. *Neuroreport*, *15*(12), 1917–1921.

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

Sloboda, J. A. (1994). What makes a musician? *EGTA Guitar Journal*, *5*, 18–22.

Smith, J. D. (1997). The place of musical novices in music science. *Music Perception*, *14*(3), 227–262.

Song, J., Skoe, E., Wong, P. C., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience*, *20*(10), 1892–1902.

Stevens, C., Keller, P., & Tyler, M. (2004). Language tonality and its effects on the perception of contour in short spoken and musical items. In S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, & P. Webster (Eds.), *Proceedings of the 8th International Conference on Music Perception and Cognition.* Adelaide, Australia: Causal Productions.

Strait, D. L., Kraus, N., Abecassis, V., & Ashley, R. (2010). Musical training's impact on neural mechanisms of selective auditory attention. In *11th International Conference on Music Perception and Cognition.* Seattle, WA.

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hearing Research*, *261*(1-2), 22–29.

Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi, R. J., & Näätänen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learning & Memory*, *8*(5), 295–300.

Trainor, L. J., Dejardins, R. N., & Rockel, C. (1999). A comparison of contour and interval processing in musicians and nonmusicians using event-related potentials. *Australian Journal of Psychology*, *51*(3), 147–153.

Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(2), 394–402.

Tramo, M. J., Cariani, P. A., Delgutte, B., & Braida, L. D. (2003). Neurobiology of harmony perception. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (p. 127-151). New York: Oxford University Press.

Trehub, S. E. (2003a). The developmental origins of musicality. *Nature Neuroscience*, *6*(7), 669-73.

Trehub, S. E. (2003b). Toward a developmental psychology of music. *Annals of the New York Academy of Sciences*, *999*, 402–413.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development*, *55*(3), 821–830.

Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: domain-general or domain-specific mechanisms? *Cognition*, *100*(1), 73–99.

Trehub, S. E., Schellenberg, E. G., & Kamenetsky, S. B. (1999). Infants' and adults' perception of scale structure. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 965–975.

Vetter, R. (1989). A retrospect on a century of gamelan tone measurements. *Ethnomusicology*, *33*(2), 217.

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The Musical Ear Test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, *20*, 188–196.

Wang, Y., Behne, D. M., Jongman, A., & Sereno, J. A. (2004). The role of linguistic experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics*, *25*(3), 449–466.

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

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

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649–58.

Wayland, R. P., & Guion, S. G. (2003). Perceptual discrimination of Thai tones by naive and experienced learners of Thai. *Cognitive Processing*, *24*(1), 113–129.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Cognitive Processing*, *54*(4), 681–712.

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

Wing, H. D. (1948). *Standardised tests of musical intelligence.* Sheffield, UK.

Wong, P. C., & Diehl, R. L. (2003). Inter- and intra-talker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research*, *46*(2), 413–422.

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

Wong, P. C., Perrachione, T. K., & Parrish, T. B. (2007). Neural characteristics of successful and less successful speech and word learning in adults. *Human Brain Mapping*, *28*, 995–1006.

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

Xu, L., & Pfingst, B. E. (2003). Relative importance of temporal envelope and fine structure in lexical-tone perception (L). *The Journal of the Acoustical Society of America*, *114*(6), 3024.

Xu, Y. (1994). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America*, *95*(4), 2240–2253.

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

Xu, Y. (1999a). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, *27*(1), 55–105.

Xu, Y. (1999b). F0 peak delay: When, where and why it occurs. In *Proceedings of the 14th International Congress of Phonetic Sciences* (p. 1881-1884).

Xu, Y. (2001). Fundamental frequency peak delay in Mandarin. *Phonetica*, *58*(1-2), 26–52.

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

Xu, Y., Krishnan, A., & Gandour, J. T. (2006). Specificity of experience-dependent pitch representation in the brainstem. *Neuroreport*, *17*(15), 1601–1605.

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, *33*(4), 319–337.

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

Yu, K. M. (2009). Contextual tonal variation in level tone languages. *The Journal of the Acoustical Society of America*, *125*(4), 2571-2571.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, *6*, 37–46.

Zentner, M. R. (1998). Infants' perception of consonance and dissonance in music. *Infant Behavior and Development*, *21*(3), 483–492.

Zongker, D. (2006). Chicken chicken chicken: Chicken chicken. *Annals of Improbable Research*, *12*(5), 16–21.

Appendix A

INSTRUMENTS

A.1 Musical Ear Test

MUSICAL EAR TEST (MET)

Test II: Comparison of melodic phrases

MELODY

Name of subject: ____________________ Date __________

EXAMPLES

	YES	NO		YES	NO
Example A	☐	☐	Example B	☐	☐

THE TEST ITSELF

	YES	NO		YES	NO		YES	NO		YES	NO
1	☐	☐	14	☐	☐	27	☐	☐	40	☐	☐
2	☐	☐	15	☐	☐	28	☐	☐	41	☐	☐
3	☐	☐	16	☐	☐	29	☐	☐	42	☐	☐
4	☐	☐	17	☐	☐	30	☐	☐	43	☐	☐
5	☐	☐	18	☐	☐	31	☐	☐	44	☐	☐
6	☐	☐	19	☐	☐	32	☐	☐	45	☐	☐
7	☐	☐	20	☐	☐	33	☐	☐	46	☐	☐
8	☐	☐	21	☐	☐	34	☐	☐	47	☐	☐
9	☐	☐	22	☐	☐	35	☐	☐	48	☐	☐
10	☐	☐	23	☐	☐	36	☐	☐	49	☐	☐
11	☐	☐	24	☐	☐	37	☐	☐	50	☐	☐
12	☐	☐	25	☐	☐	38	☐	☐	51	☐	☐
13	☐	☐	26	☐	☐	39	☐	☐	52	☐	☐

MUSICAL EAR TEST (MET)
Test I: Comparison of melodic phrases

MELODY

Subject number: ________________ Date _________

EXAMPLES

	YES	NO		YES	NO
Example A	☐	☐	Example B	☐	☐

THE TEST ITSELF

BLOCK 1	YES	NO	BLOCK 2	YES	NO	BLOCK 3	YES	NO	BLOCK 4	YES	NO
1	☐	☐	11	☐	☐	21	☐	☐	31	☐	☐
2	☐	☐	12	☐	☐	22	☐	☐	32	☐	☐
3	☐	☐	13	☐	☐	23	☐	☐	33	☐	☐
4	☐	☐	14	☐	☐	24	☐	☐	34	☐	☐
5	☐	☐	15	☐	☐	25	☐	☐	35	☐	☐
6	☐	☐	16	☐	☐	26	☐	☐	36	☐	☐
7	☐	☐	17	☐	☐	27	☐	☐	37	☐	☐
8	☐	☐	18	☐	☐	28	☐	☐	38	☐	☐
9	☐	☐	19	☐	☐	29	☐	☐	39	☐	☐
10	☐	☐	20	☐	☐	30	☐	☐	40	☐	☐

BLOCK 5	YES	NO	BLOCK 6	YES	NO	BLOCK 7	YES	NO	BLOCK 8	YES	NO
41	☐	☐	51	☐	☐	61	☐	☐	71	☐	☐
42	☐	☐	52	☐	☐	62	☐	☐	72	☐	☐
43	☐	☐	53	☐	☐	63	☐	☐	73	☐	☐
44	☐	☐	54	☐	☐	64	☐	☐	74	☐	☐
45	☐	☐	55	☐	☐	65	☐	☐	75	☐	☐
46	☐	☐	56	☐	☐	66	☐	☐	76	☐	☐
47	☐	☐	57	☐	☐	67	☐	☐	77	☐	☐
48	☐	☐	58	☐	☐	68	☐	☐	78	☐	☐
49	☐	☐	59	☐	☐	69	☐	☐			
50	☐	☐	60	☐	☐	70	☐	☐			

A.2 Tone Discrimination Test

A.2.1 Stimuli

key

syll Segmental content.

tone Tone number.

vers Which of two tokens (‘a’ or ‘b’) was used as the basis for the word.

word A full description word or the trial, respectively, including segmental content, tone, and token information.

tone ID Mean accuracy of identification of the tone by native Mandarin raters.

syll ID Mean accuracy of identification of the segmental content of the word by native Mandarin raters.

natural Mean rating of naturalness by native Mandarin raters (1 = unnatural, 5 = natural).

A.2.2 Trials

key

trial Trial number.

syll The segmental content of the first and second syllable of the trial.

tone1, tone2 The tone number of the first and second syllable, respectively.

vers1, vers2 Which of two tokens (‘a’ or ‘b’) was used as the basis for the first or second syllable, respectively.

word1, word2 A full description of the first and second word or the trial, respectively, including segmental content, tone, and token information.

contrast The tonal contrast between the two words of the trial.

syll	tone	token	word	tone ID	syll ID	natural
di	1	a	di1a	100%	100%	4.5
di	1	b	di1b	100%	100%	4
di	2	a	di2a	100%	100%	5
di	2	b	di2b	100%	100%	4
di	3	a	di3a	100%	100%	4
di	3	b	di3b	100%	50%	5
di	4	a	di4a	100%	50%	4.5
di	4	b	di4b	100%	100%	4.5
ku	1	a	ku1a	100%	100%	5
ku	1	b	ku1b	100%	100%	5
ku	2	a	ku2a	100%	100%	4
ku	2	b	ku2b	100%	100%	4
ku	3	a	ku3a	100%	100%	3.5
ku	3	b	ku3b	100%	100%	5
ku	4	a	ku4a	100%	100%	5
ku	4	b	ku4b	100%	100%	5
ma	1	a	ma1a	100%	100%	5
ma	1	b	ma1b	100%	100%	5
ma	2	a	ma2a	100%	100%	5
ma	2	b	ma2b	100%	100%	5
ma	3	a	ma3a	100%	100%	5
ma	3	b	ma3b	100%	100%	4.5
ma	4	a	ma4a	100%	100%	5
ma	4	b	ma4b	100%	100%	5

Table A.1: Identification scores and naturalness ratings by two native listeners for stimuli used in the Tone Discrimination Test used in Experiment 4 (see key).

trial	syll1	tone1	vers1	syll2	tone2	vers2	word1	word2	contrast
1	di	4	a	di	3	a	di4a	di3a	3x4
2	ku	4	a	ku	4	b	ku4a	ku4b	4x4
3	di	4	b	di	4	a	di4b	di4a	4x4
4	ma	1	a	ma	4	a	ma1a	ma4a	1x4
5	di	2	b	di	1	b	di2b	di1b	1x2
6	ma	4	b	ma	4	a	ma4b	ma4a	4x4
7	ku	3	a	ku	3	b	ku3a	ku3b	3x3
8	di	3	b	di	1	b	di3b	di1b	1x3
9	ma	3	b	ma	3	a	ma3b	ma3a	3x3
10	ku	1	b	ku	1	a	ku1b	ku1a	1x1
11	ma	4	a	ma	4	b	ma4a	ma4b	4x4
12	di	2	a	di	1	a	di2a	di1a	1x2
13	ma	4	a	ma	1	a	ma4a	ma1a	1x4
14	di	2	b	di	2	a	di2b	di2a	2x2
15	ku	3	a	ku	3	b	ku3a	ku3b	3x3
16	di	3	a	di	3	b	di3a	di3b	3x3
17	ku	3	a	ku	1	a	ku3a	ku1a	1x3
18	ma	3	b	ma	3	a	ma3b	ma3a	3x3
19	di	2	a	di	2	b	di2a	di2b	2x2
20	ma	2	b	ma	2	a	ma2b	ma2a	2x2
21	di	3	b	di	3	a	di3b	di3a	3x3
22	ma	3	b	ma	3	a	ma3b	ma3a	3x3
23	di	3	a	di	2	a	di3a	di2a	2x3
24	ku	1	a	ku	3	a	ku1a	ku3a	1x3
25	di	1	a	di	1	b	di1a	di1b	1x1
26	ma	2	b	ma	2	a	ma2b	ma2a	2x2
27	ku	4	a	ku	2	a	ku4a	ku2a	2x4
28	ma	3	a	ma	3	b	ma3a	ma3b	3x3
29	ku	3	b	ku	1	b	ku3b	ku1b	1x3
30	di	2	b	di	2	a	di2b	di2a	2x2
31	ma	4	b	ma	3	b	ma4b	ma3b	3x4
32	ku	2	a	ku	2	b	ku2a	ku2b	2x2
33	ma	2	a	ma	3	a	ma2a	ma3a	2x3
34	ku	4	a	ku	1	a	ku4a	ku1a	1x4
35	ma	1	a	ma	3	a	ma1a	ma3a	1x3
36	di	2	a	di	2	b	di2a	di2b	2x2
37	ku	4	b	ku	2	b	ku4b	ku2b	2x4
38	di	4	a	di	4	b	di4a	di4b	4x4
39	ku	2	a	ku	2	b	ku2a	ku2b	2x2
40	di	1	a	di	3	a	di1a	di3a	1x3
41	ku	2	b	ku	2	a	ku2b	ku2a	2x2
42	ma	3	a	ma	3	b	ma3a	ma3b	3x3
43	di	4	a	di	4	b	di4a	di4b	4x4
44	ma	1	b	ma	1	a	ma1b	ma1a	1x1
45	ku	3	a	ku	4	a	ku3a	ku4a	3x4
46	ma	1	b	ma	1	a	ma1b	ma1a	1x1
47	di	3	a	di	1	a	di3a	di1a	1x3
48	ku	4	b	ku	3	b	ku4b	ku3b	3x4
49	ma	4	b	ma	4	a	ma4b	ma4a	4x4
50	ku	1	a	ku	2	a	ku1a	ku2a	1x2
51	ma	3	a	ma	4	a	ma3a	ma4a	3x4
52	ku	3	a	ku	3	b	ku3a	ku3b	3x3
53	di	2	a	di	4	a	di2a	di4a	2x4
54	ma	4	a	ma	2	a	ma4a	ma2a	2x4
55	di	1	b	di	1	a	di1b	di1a	1x1
56	ma	1	b	ma	1	a	ma1b	ma1a	1x1
57	ku	1	b	ku	3	b	ku1b	ku3b	1x3
58	ma	2	a	ma	1	a	ma2a	ma1a	1x2
59	di	4	a	di	4	b	di4a	di4b	4x4
60	ku	2	b	ku	4	b	ku2b	ku4b	2x4
61	di	4	b	di	4	a	di4b	di4a	4x4
62	ku	4	b	ku	4	a	ku4b	ku4a	4x4
63	ma	1	a	ma	1	b	ma1a	ma1b	1x1
64	ku	3	b	ku	3	a	ku3b	ku3a	3x3
65	di	3	b	di	3	a	di3b	di3a	3x3
66	ma	1	a	ma	1	b	ma1a	ma1b	1x1
67	di	1	b	di	2	b	di1b	di2b	1x2
68	ma	2	b	ma	2	a	ma2b	ma2a	2x2
69	di	2	b	di	3	b	di2b	di3b	2x3
70	ku	2	a	ku	3	a	ku2a	ku3a	2x3
71	di	4	b	di	2	b	di4b	di2b	2x4
72	ku	3	b	ku	2	b	ku3b	ku2b	2x3

Table A.2: Trial list for Tone Discrimination Test used in Experiment 4 (part 1, trials 1–72; see key).

trial	syll1	tone1	vers1	syll2	tone2	vers2	word1	word2	contrast
73	di	1	a	di	1	b	di1a	di1b	1x1
74	ma	2	a	ma	4	a	ma2a	ma4a	2x4
75	di	1	b	di	1	a	di1b	di1a	1x1
76	ma	2	b	ma	1	b	ma2b	ma1b	1x2
77	di	4	b	di	4	a	di4b	di4a	4x4
78	ku	1	b	ku	4	b	ku1b	ku4b	1x4
79	di	3	b	di	2	b	di3b	di2b	2x3
80	ku	2	b	ku	1	b	ku2b	ku1b	1x2
81	di	2	a	di	2	b	di2a	di2b	2x2
82	ku	3	a	ku	2	a	ku3a	ku2a	2x3
83	di	3	a	di	3	b	di3a	di3b	3x3
84	ma	1	a	ma	2	a	ma1a	ma2a	1x2
85	di	1	a	di	1	b	di1a	di1b	1x1
86	ku	3	b	ku	3	a	ku3b	ku3a	3x3
87	di	1	a	di	4	a	di1a	di4a	1x4
88	ku	4	b	ku	4	a	ku4b	ku4a	4x4
89	di	1	b	di	1	a	di1b	di1a	1x1
90	ku	3	b	ku	3	a	ku3b	ku3a	3x3
91	di	1	b	di	4	b	di1b	di4b	1x4
92	ma	3	b	ma	2	b	ma3b	ma2b	2x3
93	di	2	b	di	2	a	di2b	di2a	2x2
94	ma	1	b	ma	3	b	ma1b	ma3b	1x3
95	di	4	b	di	1	b	di4b	di1b	1x4
96	ma	4	b	ma	2	b	ma4b	ma2b	2x4
97	di	4	a	di	1	a	di4a	di1a	1x4
98	ma	3	a	ma	2	a	ma3a	ma2a	2x3
99	di	3	b	di	4	b	di3b	di4b	3x4
100	ku	3	b	ku	4	b	ku3b	ku4b	3x4
101	di	3	a	di	3	b	di3a	di3b	3x3
102	ku	2	b	ku	2	a	ku2b	ku2a	2x2
103	di	4	b	di	3	b	di4b	di3b	3x4
104	ku	2	b	ku	2	a	ku2b	ku2a	2x2
105	ma	2	a	ma	2	b	ma2a	ma2b	2x2
106	ku	1	a	ku	1	b	ku1a	ku1b	1x1
107	ma	4	a	ma	4	b	ma4a	ma4b	4x4
108	ku	1	b	ku	1	a	ku1b	ku1a	1x1
109	ma	2	b	ma	3	b	ma2b	ma3b	2x3
110	di	2	a	di	3	a	di2a	di3a	2x3
111	ku	2	a	ku	4	a	ku2a	ku4a	2x4
112	ma	3	a	ma	3	b	ma3a	ma3b	3x3
113	di	2	b	di	4	b	di2b	di4b	2x4
114	ma	1	b	ma	2	b	ma1b	ma2b	1x2
115	ku	2	b	ku	3	b	ku2b	ku3b	2x3
116	di	1	a	di	2	a	di1a	di2a	1x2
117	ku	1	a	ku	4	a	ku1a	ku4a	1x4
118	ma	2	a	ma	2	b	ma2a	ma2b	2x2
119	ku	2	a	ku	2	b	ku2a	ku2b	2x2
120	ma	1	b	ma	4	b	ma1b	ma4b	1x4
121	ku	4	a	ku	4	b	ku4a	ku4b	4x4
122	ma	3	b	ma	1	b	ma3b	ma1b	1x3
123	ku	4	b	ku	4	a	ku4b	ku4a	4x4
124	ma	3	a	ma	1	a	ma3a	ma1a	1x3
125	ku	1	b	ku	1	a	ku1b	ku1a	1x1
126	di	3	a	di	4	a	di3a	di4a	3x4
127	ma	3	b	ma	4	b	ma3b	ma4b	3x4
128	ku	1	b	ku	2	b	ku1b	ku2b	1x2
129	ma	2	b	ma	4	b	ma2b	ma4b	2x4
130	ku	2	a	ku	1	a	ku2a	ku1a	1x2
131	ma	4	b	ma	1	b	ma4b	ma1b	1x4
132	ku	4	b	ku	1	b	ku4b	ku1b	1x4
133	ma	4	a	ma	4	b	ma4a	ma4b	4x4
134	ku	1	a	ku	1	b	ku1a	ku1b	1x1
135	ma	4	a	ma	3	a	ma4a	ma3a	3x4
136	di	4	a	di	2	a	di4a	di2a	2x4
137	ku	4	a	ku	4	b	ku4a	ku4b	4x4
138	ma	2	a	ma	2	b	ma2a	ma2b	2x2
139	ku	4	a	ku	3	a	ku4a	ku3a	3x4
140	ma	4	b	ma	4	a	ma4b	ma4a	4x4
141	di	1	b	di	3	b	di1b	di3b	1x3
142	ma	1	a	ma	1	b	ma1a	ma1b	1x1
143	di	3	b	di	3	a	di3b	di3a	3x3
144	ku	1	a	ku	1	b	ku1a	ku1b	1x1

Table A.3: Trial list for Tone Discrimination Test used in Experiment 4 (part 2, trials 73–144; see key).

TONE TEST

☐

Subject number: ____________________   Date __________

EXAMPLES

	YES	NO		YES	NO
Example A	☐	☐	Example B	☐	☐

THE TEST ITSELF

BLOCK 1			BLOCK 2			BLOCK 3			BLOCK 4		
	YES	NO		YES	NO		YES	NO		YES	NO
1	☐	☐	11	☐	☐	21	☐	☐	31	☐	☐
2	☐	☐	12	☐	☐	22	☐	☐	32	☐	☐
3	☐	☐	13	☐	☐	23	☐	☐	33	☐	☐
4	☐	☐	14	☐	☐	24	☐	☐	34	☐	☐
5	☐	☐	15	☐	☐	25	☐	☐	35	☐	☐
6	☐	☐	16	☐	☐	26	☐	☐	36	☐	☐
7	☐	☐	17	☐	☐	27	☐	☐	37	☐	☐
8	☐	☐	18	☐	☐	28	☐	☐	38	☐	☐
9	☐	☐	19	☐	☐	29	☐	☐	39	☐	☐
10	☐	☐	20	☐	☐	30	☐	☐	40	☐	☐

BLOCK 5			BLOCK 6			BLOCK 7			BLOCK 8		
	YES	NO		YES	NO		YES	NO		YES	NO
41	☐	☐	51	☐	☐	61	☐	☐	71	☐	☐
42	☐	☐	52	☐	☐	62	☐	☐	72	☐	☐
43	☐	☐	53	☐	☐	63	☐	☐	73	☐	☐
44	☐	☐	54	☐	☐	64	☐	☐	74	☐	☐
45	☐	☐	55	☐	☐	65	☐	☐	75	☐	☐
46	☐	☐	56	☐	☐	66	☐	☐	76	☐	☐
47	☐	☐	57	☐	☐	67	☐	☐	77	☐	☐
48	☐	☐	58	☐	☐	68	☐	☐	78	☐	☐
49	☐	☐	59	☐	☐	69	☐	☐	79	☐	☐
50	☐	☐	60	☐	☐	70	☐	☐	80	☐	☐

BLOCK 9

	YES	NO
81	☐	☐
82	☐	☐
83	☐	☐
84	☐	☐
85	☐	☐
86	☐	☐
87	☐	☐
88	☐	☐
89	☐	☐
90	☐	☐

BLOCK 10

	YES	NO
91	☐	☐
92	☐	☐
93	☐	☐
94	☐	☐
95	☐	☐
96	☐	☐
97	☐	☐
98	☐	☐
99	☐	☐
100	☐	☐

BLOCK 11

	YES	NO
101	☐	☐
102	☐	☐
103	☐	☐
104	☐	☐
105	☐	☐
106	☐	☐
107	☐	☐
108	☐	☐
109	☐	☐
110	☐	☐

BLOCK 12

	YES	NO
111	☐	☐
112	☐	☐
113	☐	☐
114	☐	☐
115	☐	☐
116	☐	☐
117	☐	☐
118	☐	☐
119	☐	☐
120	☐	☐

BLOCK 13

	YES	NO
121	☐	☐
122	☐	☐
123	☐	☐
124	☐	☐
125	☐	☐
126	☐	☐
127	☐	☐
128	☐	☐
129	☐	☐
130	☐	☐

BLOCK 14

	YES	NO
131	☐	☐
132	☐	☐
133	☐	☐
134	☐	☐
135	☐	☐
136	☐	☐
137	☐	☐
138	☐	☐
139	☐	☐
140	☐	☐

BLOCK 15

	YES	NO
141	☐	☐
142	☐	☐
143	☐	☐
144	☐	☐

Appendix B

FORMS

Date:________________ Location:________________

Consent to Participate in Research

Study Title: Crosslinguistic Perception of Pitch in Language and Music
Principal Investigator: Evan D. Bradley, Dept. of Linguistics & Cognitive Science
Advisors: Drs. Jeffrey Heinz & Irene Vogel

1 Purpose and Description of Study

Purpose

- The purpose of this study is to investigate how the language you speak affects how you perceive sounds, how you learn new languages, and how you perceive music.
- Approximately 300 participants will take part in this study.
- You volunteered or were asked to participate in this study because you meet the following criteria:
  - You are a native speaker of either English, Mandarin Chinese, or Yoruba, or a native speaker of both Spanish and English.
  - You are over 18 years of age.

Description

- Your participation in the study will between 1 and 8 hours; you may be asked to complete your participation in a single session (up to one hour), or in multiple sessions.
- You will be offered frequent opportunities for rest and refreshment during your participation.
- You will take part in perceptual and memory tests, in which you will listen to language sounds and musical pieces, and respond either via a computer or with pencil and paper.
- **Please note:** These tests are not intended to reflect or predict past or future academic, linguistic, or musical achievement or aptitude; nor are they intended to diagnose any illness or learning disability.
- **Important:** If you are participating in this study as part of a class, your grades and test scores may be included in the data. These grades will not be shared with anyone but the Investigator, and will be kept confidential along with the rest of your data.

  Initials: _______ I understand that my grades and test scores may be included in the data for this study. I further understand that they will be kept confidential, and will not be shared with anyone other than the investigator.

2 Conditions of Participation

- You may choose to end your participation in this study at any time without consequence.
- The investigator may terminate your participation at any time if it is determined that you do not meet eligibility requirements, or are unable to comply with study procedures.
- If you wish, you may be notified about publications resulting from this study.

Confidentiality

- Your identity will be known only to the Investigator, who will maintain your paper and electronic records securely.
- Data associated with your participation, including your grades and test scores, will be identified only with a numerical code.
- Your personal information may be maintained for up to five years, after which it will be securely destroyed.
- Data which does not contain information which can personally identify you may be retained indefinitely for future analyses.
- Results of this study will be published, but will not include information that can personally identify you.

1

3 Risks & Benefits

- There are no known risks of participating in this study.
- There are no immediate benefits to you resulting from participation in this study, but the completion of this study will contribute to knowledge about perception and learning, and is expected contribute to methods used in the teaching of language and music.

4 Compensation

- If you are participating in this study as a part of a class, you will receive a small amount of credit in compensation for your participation. Your instructor will offer you an alternative assignment which you may complete for credit instead of participating in this study.
- If you are not participating as a part of a class, you will be paid $10 per hour of participation ($5 for each half hour or portion thereof), to be collected at the conclusion of your participation.
- If you voluntarily withdraw from the study, or if the investigator terminates your participation, you will still be compensated for the time you have participated.

5 Contacts

- If you have any questions about this research, contact Evan Bradley at `yevb@udel.edu` or `(302) 533-8094`.
- If you have questions about your rights as a study participant, you may contact the Human Subjects Review Board, University of Delaware Research Office at `hsrb-research@udel.edu` or `(302) 831-2137`.

6 Participant Information

Name (print): ____

Email address: ____

Phone number: ____

Native language: ____

Date of Birth (month/day/year): ______/______/______

7 Consent

I have read and understood this consent form and I voluntarily agree to participate in the study described; I understand that I may withdraw from the study at any time.

Signature: ____ ______/______/______

Investigator: ____ ______/______/______

Subject #:________ Date:________ Location:________

Language Profile

Demographic Information

Date of birth (mm/dd/yyyy): ____/____/____ **Sex:** ________

Level of education: ________

Occupation: ________

Residence History

Place of birth (city, country): ________

Current residence (city, country): ________

What other countries you have lived/studied in, and for how long? ________

Language History

Have you ever been diagnosed with a learning disability? ________

Have you ever been diagnosed with a hearing or vision problem? ________

Native language(s): ________

Please list all of the languages you can speak, and how well you can speak them:

1. ________
2. ________
3. ________
4. ________

Are there any languages you can understand, but not speak? if so, how well?

1. ________
2. ________
3. ________

When you were a child, what language(s) were spoken in your home?

1. ________
2. ________
3. ________

What language(s) were spoken in the community/school where you grew up?

1. ________
2. ________
3. ________

Did you study any languages in school? If so, for how long?

1. ________
2. ________
3. ________

Subject #:__________ Date:__________ Location:__________

Music Profile

Performance

Can you sing or play any musical instruments? If so, please indicate:

- which instrument/type of singing
- how long you have played/sung
- your estimated skill level on a scale of 1–10 (1 = beginner, 10 = professional)

1. __________
2. __________
3. __________
4. __________

Have you taken any kind of music lessons? If so, what kind? and for how long?

1. __________
2. __________
3. __________
4. __________

Have you taught yourself to sing or play any instruments? If so, which? and for how long?

1. __________
2. __________
3. __________
4. __________

Have you ever sung in a choir? If so, what kind, and for how long? __________

____

How good would you say you are at singing? (1 = terrible, 10 = professional) __________

Have you ever studied music theory or composition? __________

Can you read music? If so, how well? __________

Listening

What are your favorite kinds of music to listen to? __________

____

____

How many hours per day do you listen to music? __________

Would you say you have a "good ear" for music? (1 = tone deaf, 10 = very good) __________

Do you listen to any of the following kinds of music? How often?

American/Western Pop music? __________
European/Western Classical music? __________
Chinese/Asian Pop music? __________
Traditional Chinese music? __________
African Pop music? __________
Traditional African music? __________
Indian Classical/Traditional music? __________

UNIVERSITY OF DELAWARE

RESEARCH OFFICE

210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE:	July 11, 2011
TO:	Evan D. Bradley, MA
FROM:	University of Delaware IRB
STUDY TITLE:	[238566-1] Crosslinguistic Perception of Pitch in Language and Music
SUBMISSION TYPE:	New Project
ACTION:	APPROVED
APPROVAL DATE:	July 11, 2011
EXPIRATION DATE:	July 10, 2012
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # 5, 7

Thank you for your submission of New Project materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

COLOPHON

This document was composed in LaTeX, based on the `UDThesis` package maintained by Anita Schwartz—whose assistance with technical issues was appreciated—with modifications by the author. Other useful packages included:

- apacite
- appendix
- attrib
- hyperref
- multirow
- multicol
- musixtex
- nameref
- pdfpages
- setspace
- tipa

LICENSE & COPYRIGHT