Hmong Music and Language Cognition: An Interdisciplinary Investigation

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By
Nicholas Frederick Poss, M.A.
Graduate Program in Music

The Ohio State University
2012

Dissertation Committee:
Udo Will, Advisor
Mark Pitt
Ryan Skinner
Copyright by

Nicholas Frederick Poss

2012
Abstract

Speech surrogacy, which includes the performance of verbal messages on musical instruments, is found in a variety of cultures. The developing field of music and language cognition can benefit from the study of these communicative forms, which confound our expectations of the boundaries between speech and music. Previous studies have focused on semiotic relationships of similarity between musical sound and speech. While this type of analysis can suggest strategies for decoding messages, it cannot explain how listeners make use of this information. Using methodology derived from psycholinguistics, this dissertation investigates speech surrogate cognition from the perspective of Hmong culture to find out how listeners understand verbal messages encoded in performances on aerophones called *raj*.

In one experiment, musical phrases of varying lengths were presented to skilled listeners to investigate the strategies used in understanding performances. The results show that listeners are highly successful at identifying phrases. For ambiguous words, listeners relied mainly on the established relationships between musical pitch and lexical tone to infer meaning rather than broad distinctions between types of syllable onsets. This demonstrates a problem with the semiotic approach to analyzing speech surrogates: listeners do not necessarily make use of
everything encoded in the signal. Finally, there were different response patterns for phrases of different lengths, indicating that the context of messages affects how listeners interpret them.

The second experiment looked at the effect of individual pitches on lexical selection to see if instrumental sounds might speed access to relevant lexical items. An auditory priming paradigm was used in which subjects had to repeat words and pseudowords in the Hmong language. These target items were primed by musical sounds and pitch contours derived from spoken words. The primes either matched or mismatched the targets and were compared against white noise in the envelope filter of a spoken syllable to examine the effect on reaction time. It was hypothesized that words primed by matched musical sounds and pitch contours would be repeated more quickly than words primed by mismatched sounds if the relationship of similarity acted upon the mental lexicon. The results showed no effect for matched or mismatched primes but did show an effect for primes containing pitch content versus the control. Pitched primes speeded reaction times to both words and pseudowords, suggesting that the effect is not related to lexical processing. It is possible that the pitched sounds primed areas of the pre-motor cortex, which is involved in planning movements of the vocal tract, resulting in a faster response. This effect was also found in subsequent experiments with speakers of another tonal language (Mandarin) who do not practice speech surrogacy.
This research demonstrates the benefit of interdisciplinary research that includes an ethnographic approach. Music and speech are not so neatly categorized in many cultures and by studying phenomena on this border we will better understand not only speech surrogate cognition, but also the role of pitch in language cognition as well as the effect of personal experience on cognition.
Dedicated to my son
Acknowledgments

This project would not have been possible without support and expertise of several people and organizations. My research was funded by Ethnomusicology Field Research Grants from The Ohio State University in 2006, 2007, and 2008. I also received the Jack and Zoe Johnstone Graduate Award for Excellence in Musicology in 2008.

Lopao Vang (Vaj Tooj Pov) lent his voice to the recordings that were used in the priming experiment and was instrumental in recruiting subjects in the Madison, WI area. In the Twin Cities, I relied heavily on the help of the Hmong Cultural Center and the Center for Hmong Arts and Talent. Kathy Mouacheupao and Tou Saiko Lee were especially helpful in finding participants and allowing me to use their offices to conduct experiments. I am also deeply indebted to the many people who took part in the experiments and the musicians who were so generous with the time.

My advisor, Udo Will, has been a constant source of encouragement was always willing to help me push through the obstacles I encountered. I am grateful to my mother and father who have been unwavering in their support of my work and
who have contributed greatly to the completion of this document. Most of all I
would like to thank my wife who believed in me even when I did not and whose
patience and love have seen me through.
Vita

2001.................................................................B.M. Music Theory, University of Wisconsin-Eau Claire

2002 to 2006..................................................Graduate Teaching Associate, The Ohio State University

2005.................................................................M.A. Ethnomusicology, The Ohio State University

2009 to present ..............................................Lecturer, The Ohio State University

2010 to present ..............................................Adjunct Music Professor, Ashland University

Publications


Fields of Study

Major Field: Music
# Table of Contents

Abstract .......................................................................................................................... ii  
Dedicated to my son....................................................................................................... v  
Acknowledgments ......................................................................................................... vi  
Vita .................................................................................................................................. viii  
List of Tables .................................................................................................................... xiii  
List of Figures ................................................................................................................... xiv  

Chapter 1: Introduction ..................................................................................................... 1  
1.1: Speaking musically ................................................................................................. 1  
1.2: Musical speech surrogacy ....................................................................................... 7  
1.3: Studying musical speech surrogates with experiments ......................................... 11  
1.4: The study of music and language ........................................................................... 16  

Chapter 2: Hmong speech surrogacy ................................................................................. 22  
2.1: Location of the study ............................................................................................. 22  
2.2: The history of Hmong-Americans .......................................................................... 23  
2.3: Previous research on Hmong music ....................................................................... 26
Chapter 3: Understanding phrases on the raj

3.1: Introduction

3.2: Subjects

3.3: Experimental materials

3.4: Results

3.5: Discussion

Chapter 4: Priming the mental lexicon with pitch contours

4.1: Introduction

4.2: Materials

4.3: Subjects

4.4: Procedure

4.5: Results

4.6: Discussion

Chapter 5: New directions in research

5.1: The cross-cultural study of pitch priming

5.2: Formulaic language on the Hmong raj

5.3: Problems and possibilities in interdisciplinary research

5.4: Conclusion
List of Tables

Table 2.1. The mapping of lexical tones onto pitches in *suab raj nplaim*.........................36

Table 2.2. Consonants and their associated articulation types in *tshuab raj*.....................41

Table 3.1. Instruments played by the 13 subjects who claimed to understand
messages in traditional Hmong music.................................................................50

Table 3.2. Reported musical practices of the subjects.....................................................51

Table 3.3. The words of the phrases used in the experiment, as described by the
original performer, and an English translation......................................................54

Table 3.4. The number of requests for repetition of the experimental items. ...............56

Table 3.5. Subject responses to the 16 items in order of presentation. .........................58

Table 3.6. Words reported by the subjects that did not match those reported by the
performer that did match in tone and onset, just tone, or just onset.......................62
List of Figures

Figure 2.1. Picture of Vang Tong Thor playing the *raj nplaim* with a section from a plastic bottle attached as a resonator. Madison, Wisconsin, 2005. ........................................29

Figure 2.2. Transcription of a *nqes* performed in *suab raj nplaim* from a longer performance by Sai Pao Lee recorded in 2005, Minneapolis, Minnesota.........................42

Figure 3.1. The percentage of words reported by the subjects that matched those reported by the performer for each item.........................................................................................................60

Figure 4.1. Instrumental (left) and vocal (right) primes with melodic contours corresponding to the four Hmong tones used in the experiment. .................................85

Figure 4.2. The organization of the experimental materials.............................................87

Figure 4.3. Response times for the control, match, and non-match conditions for pseudowords and words........................................................................................................89

Figure 4.4. Facilitation in the match condition (RTco – RTma) for both instrumental and vocal primes split by pseudowords and words.........................................................92

Figure 4.5. Facilitation in the non-match condition (RTco – RTma) for both instrumental and vocal primes split by pseudowords and words.............................................93

Figure 5.1. Interaction plot for lexical status and prime condition. p: pseudowords, w: words, c: control, m: matching, n: nonmatching, rt: reaction time.................................101
Figure 5.2. Interaction plot for the effects of lexical status (w: words; p: pseudowords) and prime condition (c: control; m: matching contour) on reaction time.
Chapter 1: Introduction

1.1: Speaking musically

In the Hmong language there is no clear distinction between music and speech. When you speak to someone else, you “hais lus” (“speak words”).¹ When you sing traditional extemporaneous poetry called kwv txhiaj, you “hais kwv txhiaj” (literally, “speak poetry”). When you play a melody on a traditional flute or free-reed pipe, you “tshuab hais lus” (literally, “blow speak words”). The melodies performed on instruments convey more than emotions or familiar topics, they convey words. In American culture, we might hear the melody to a familiar tune, like “Happy Birthday,” and immediately think of familiar lyrics. But Hmong musicians are not limited to a repertoire of songs—they can put together words and melodic phrases on the spot to respond to the current situation.

I encountered this practice in 2000 while producing an educational video about traditional Hmong music in Eau Claire, Wisconsin. Many Hmong people supported the American side in the Secret War, fought in Laos during the 1960s and 70s. When the communist forces took control of the government there in 1975, Hmong people

¹ The Hmong words are written in Romanized Popular Alphabet (RPA), a system for writing the Hmong Der and Mong Leng dialects developed in the 1950s. Most words follow a consonant-vowel
were singled out for retribution. This led to a mass exodus of refugees across the Mekong River into Thailand where people sometimes waited several years to be resettled in a third country. Over the last 35 years, tens of thousands of Hmong people came to the United States, the last group arriving in 2005. Today, 260,076 Hmong people live in the U. S. (Census 2010).

As a child in the 1980s, I went to school with newly arrived Hmong refugees, and while I knew them as friends and classmates, I learned little about their culture or history. Several years later, attending the University of Wisconsin-Eau Claire, I had the opportunity to produce a video about Hmong music to be used by educators in the classroom. It was then that I met highly skilled musicians like Joe Bee Xiong and his cousin Tou Ger Xiong, among others. They patiently answered my questions, recorded interviews, and provided performances for the video. Tou Ger Xiong was particularly well known for his ability on the raj, a class of traditional aerophones that includes flutes and free-reed pipes.\(^2\) He played melody after melody and explained in English the Hmong words he had communicated. I asked him how he did it. “Hmong music is not like normal music,” he said. “We play word by word into the flute.” He demonstrated by speaking a few words and then playing them on the flute, but I could not make the connection. I was still left wondering, how do they do

\(^2\) Raj is pronounced /ˈtæ/ with a high-falling tone. There is no exact equivalent, but “dra” approximates the pronunciation.
it? How do musicians transform verbal phrases into musical ones? How do listeners understand these musical phrases as words?

For several years, I visited performers in Minnesota and Wisconsin learning about their musical histories and performance practices. I asked them questions about how they made the connection between musical sounds and words. I also recorded performances to analyze later. Some of the musicians even taught me how to play a few phrases on the free-reed pipe called *raj nplaim.* Through this ethnographic approach, I was able to learn much about the Hmong musical system and how the *raj* and speech surrogacy fits into Hmong-American life (Poss 2005).

When I asked Hmong musicians how they understood words played on the *raj,* I received a variety of responses. Some said that if you could speak the Hmong language then you could understand the words. Yet many Hmong people I met, especially young Hmong Americans who had grown up in the United States, were unable to understand the words in *raj* performances. Tou Ger Xiong explained to me that the *suab* of the word has to match the *suab* of the *raj.* *Suab* means “sound” and can be used in variety of ways. Hmong is a tonal language and *suab* can refer to the lexical tones of words. It can also refer to the pitches played on a musical instrument. He spoke several words and played the corresponding notes on the instrument, but he could not explain what the nature of the match was. Previous

---

3 *Raj* means “tube” and *nplaim* refers to the metal reed found in instruments like the *qeej,* a free-reed mouth organ, and the free-reed pipe, *raj nplaim.* The approximate English pronunciation is “dra nblai.”
studies by Eric Mareschal (1976) and Amy Catlin (1982) demonstrated that there are consistent relationships between the lexical tones of Hmong words and the pitch levels within a given range in a musical performance. This was verified in my analysis, as well. Performers use a variety of scale-types, but within these types, lexical tone-to-musical pitch relationships remain the same. For example, a high-tone word might be consistently played on the highest pitch in the range, while a falling-tone word might be played the lowest pitch.  

These connections between lexical tone and musical pitch are not sufficient to communicate words. A typical pitch range only has four or five notes—not enough to differentiate between the seven lexical tones of the Hmong language that are commonly played on the raj, let alone distinguish individual words. My research also showed that performers draw on a limited repertoire of topics and formulaic sayings. While some long phrases are common among players, most performances are built out of three- to four-word expressions that can be altered and combined in new ways to address different situations. For example, different names and pronouns might be inserted or the order of the ideas might be changed. This allows listeners to predict the sayings that a performer might use. The resulting melodies sound similar to those in traditional vocal genres, but the verbal content is organized differently. In one of the most common vocal genres, kwv txhiaj, singers

---

4 A detailed description of these systems can be found Chapter 2
5 I use the term pitch range to refer to the group of pitches that are played within the context of a single suab, or style, on the raj.
adhere to a pattern of rhyming words. In *tshuab raj* ("playing the *raj"*) there is no rhyme scheme. Also, phrases are repeated throughout the performance, a technique found in other musical practices that communicate words through the sound of instruments, usually referred to as speech surrogates. Such redundancy provides more opportunities for listeners to figure out the underlying message (Ong 1977:424).

This ethnographic approach was effective in describing the way the *raj* is played and how it fits in to Hmong-American society, but it cannot fully answer my original questions. Some of the lexical tone-to-musical pitch mappings appear to be based on relationships of acoustic similarity. For example, in one pitch range, high-tone words were played on high pitches and low-tone words were played on low pitches. From this evidence, it is tempting to think Hmong people play and understand musical messages based on perceived similarities in the sounds of words and music. But the mappings of other lexical tones and pitch levels are not based on an obvious relationship of similarity. In my research I found one pitch range where falling-tone words were played on the lowest pitch while in another pitch range they were played on the highest (2005:118-125. In another case, two lexical tones were mapped to two different pitches based on what lexical tone preceded them, not unlike tone sandhi in speech. But these changes in pitch level did not correspond to tone sandhi in spoken Hmong for those lexical tones (141-143).

Whether or not the sounds of music and speech are perceptually similar, the question remains: how do listeners make use of the signal to understand the
message? This led to the development of two experiments to explore various possibilities. In one experiment, skilled listeners were played phrases of various lengths from raj performances and asked to report the words they understood. This was done to address a number of related questions: To what extent do listeners understand the words of tshuab raj? How does understanding relate to the length of the phrase? What strategies do listeners use to interpret ambiguous phrases? For example, would listeners make use of the lexical tone-to-pitch mappings when coming up with words? Would they make use of broad distinctions in syllable-onset types indicated by different types of articulations on the raj? Would they rely on familiar phrases or would they generate novel expressions to fit the features of the musical sound?

The second experiment drew on methodology from psycholinguistics to find out if the pitch of a musical sound or a lexical tone could bring to mind words with a related lexical tone. The lexical tone-to-pitch mappings found in Hmong music are not arbitrary, which suggests that there might a perceptual link between music and speech in Hmong culture. This link could provide an advantage for bringing to mind words with lexical tones related to the melody or it could inhibit the activation of incompatible words. In the experiment, subjects repeated (shadowed) target words and pseudowords that had been primed with a variety of pitched sounds. For example, would listeners respond more quickly to word that had been primed by a sound from a raj that matched how it would be played on the instrument?
Performing these experiments posed special challenges. First, the Hmong-American community is widely dispersed across Minnesota and Wisconsin. While I had contacts in cities with fairly large communities, like the Twin Cities and Madison, Wisconsin, it would not be feasible to set up a central laboratory. Rather, I developed the experiments to be portable so that they could be performed wherever I met someone who was willing to participate. While the second experiment was intended for Hmong-Americans regardless of their musical experience, the first experiment required subjects who understood *tshuab raj*. My previous research found that the practice was found almost exclusively among Hmong-Americans who had learned it Laos or Thailand before resettling in the United States. They tend to be older people, many of whom speak little or no English. This made subject recruitment a challenge and heightened the need for a portable system that could be used to run the experiment. I was able to rely on some contacts to find participants via word of mouth and received help from the Hmong Cultural Center in St. Paul, Minnesota where I was able to test several people who were making use of the services there. As a result, I was able to find enough subjects to perform both experiments.

1.2: Musical speech surrogacy

In the words of Theodore Stern, speech surrogates are the “conversion of human speech into equivalent sounds for transmission in signaling systems” (1957). He goes on to discuss drum and whistle languages as well as surrogates played on other
musical instruments. But under this very broad definition, systems like Morse code, flag semaphore, secret speech, and other types of signals that share little in common could also be included. Julian Meyer has used the term “talking musical instruments,” which provides a clearer set of boundaries (2004). But it excludes whistle languages that are not performed on instruments. The term “drum and whistle languages” has been used by several authors (e.g. Sebeok & Umiker-Sebeok 1977) even when describing systems that are neither drummed nor whistled but rather played on musical instruments (e.g. Nöth 1995; Meyer 2005). Again, the terminology is unclear and limiting.

Yet drum and whistle languages and talking musical instruments share an important feature: the conversion of speech into a musical medium. The signal is based on pitch, rhythm, timbre, and other characteristics shared by speech and music, but it is performed by musical means—drums, whistling, flutes, etc. Perhaps the best term to describe these related phenomena is “musical speech surrogates,” which has been used by Stephen Feld and Aaron Fox (1994) among others. Even this term highlights a significant cultural bias. For many of the cultures studied under this lens, speech and music are not so easily partitioned. As noted above, Hmong language uses the same words for speaking and playing words on musical instruments. In fact, all traditional Hmong music communicates words, except for the ostinatos played by musicians who perform acrobatic dances at New Year festivals while playing the

---

6 Meyers himself differentiates between whistled language and talking musical instruments, even when they are related in a cultural practice.
free-reed mouth organ, qeej. When someone plays a traditional Hmong flute, for example, it is assumed that they are speaking words. For some people, this might be generalized to other music. One older Hmong woman who could play tshuab raj told me that she enjoyed listening to Mozart’s symphonies but that she could not understand the words. These fuzzy boundaries of musical speech surrogates provide opportunities for understanding how music and language are related culturally and cognitively.

Most studies of speech surrogates examine them from the standpoint of the relationship between the sound of speech and the sound of the surrogate signal. This can be found even in very early studies like those of Joseph Lajard, which were reported by M. F. Force and J. N. B. Hewitt in an 1893 edition of American Anthropologist. Lajard was the first to recognize that the whistle language of La Gomera in the Canary Islands was based on the sound of spoken Spanish. He contrasted this with the whistle languages of "cultured communities" like carpenters, stone-masons, and the "poaching, marauding, and ex-convict gentry" of Paris. These languages were based on "conventional signs" and could not reproduce all spoken language.

Later studies were able to pinpoint important similarities between the sound of speech and the sound of the whistle. In a study of Silbo Gomero, Rialland (2005) found that the second formants of the five vowels whistled (/i e a o u/) correspond to the first harmonic of the whistle—maintaining the hierarchy of pitch levels between the vowels. She also found parallel pitch contours in speech and whistling
during vowel-consonant or consonant-vowel transitions. For example, the second formant typically rises as the vowel /o/ approaches a dental consonant and this same rise can be found in the first harmonic of the whistled version of /ot/. There are broader similarities, as well, like the overall shape of the amplitude envelope. Riall and also notes important differences, some of which can be explained by the limits of whistling. In the analysis of one whistler, she found that the first harmonics of the whistled vowels /i/, /e/, and /a/ were very close in pitch to the second formants of the spoken versions of those vowels. The first harmonics of the whistled vowels /o/ and /u/, however, were much higher and closer in pitch space than the second formant of their spoken counterparts. She attributes this to a pitch floor for whistling that is about 400 Hz above the lowest second formant of a spoken vowel.

In some musical speech surrogates, there is very little in common between the sound of speech and the musical sound. According to Kenelm O. L. Burridge (1959), the Tangu people of New Guinea communicate messages on slit gongs by playing melodies associated with specific phrases. These melodies have little in common with the sound of the words, although in some cases there is a similarity in the rhythms. Burridge refers to these musical symbols as "call-signs," which are assigned to people, animals, places, and common events (1238). This wide variation in the encoding of speech signal into a musical form, from individual vowels to complete phrases, demonstrates the importance of the context in communication.

The nature of the relationship between sound and signal was most thoroughly explored by Theodore Stern who came up with a classification of speech surrogates
based on semiotic theory pioneered by Charles Peirce (1934). According to Peirce, a sign can denote its object in three ways: an icon resembles its object, an index points to its object, and a symbol is associated with its object by convention. Stern applies these different types of representation to different levels within oral messages communicated by speech surrogates: phonemic, lexical/morphological, and message unit. He notes that most speech surrogates are based on “abridgement” in which “each transmitted sign exhibits significant resemblance to a corresponding sound of the base message” at the phonemic level (1957:488). In tshuab raj, this iconic relationship is evident when performers play words with high tones on high pitches within a given range. Stern contrast abridgement with “encoding,” a symbolic relationship in which there is no similarity between the message and signal at the phonemic level. On the lexical or message unit level, Stern uses the term “ideograph” to refer to signals that represent the message without any relationship of resemblance. He goes on to discuss a variety of speech surrogates from around the world and how they make use of abridgement, encoding, and ideographs.

1.3: Studying musical speech surrogates with experiments

Stern notes that “[t]he labeling of a system as encoding, abridging, or ideographic must accordingly be approached with circumspection, and it should rest upon an adequate knowledge of the base language” (494). But even with a deep knowledge of the language, such classifications are based on assumptions. Just because a relationship of similarity can be found between the message and the signal does not
necessarily mean that a listener makes use of this in understanding the words. Yet few studies of speech surrogates have ventured beyond this level of description. One of the first researchers to go beyond semiotic description was René-Guy Busnel who performed experiments among the practitioners of a whistle language in Kusköy, Turkey (1970). While the tests did not investigate what features of the whistle listeners used to understand the words, he did demonstrate that people are highly successful in interpreting messages even in unusual experimental situations.

Busnel began with recordings of ten whistlers and then broke down their performances into 45 words or short phrases. Subjects listened to the fragments in random order and identified isolated words or short phrases correctly 70% of the time. Besides using whistles from 10 different whistlers, the test was made more difficult by selecting whistled words that formed minimal pairs (e.g. yapici and kapici). According to Busnel, the subjects found the experiment very unusual, which made them ill at ease and tense. Five men and two women were tested, but the women were especially uncomfortable because the test givers were men and they were Muslim women. Their responses were unusually error prone and were not included in the overall totals. In a follow up experiment, Busnel took two men and had them take turns whistling numbers to each other. There were ten trials of ten items each. The trials consisted of random numbers, but one consisted of large numbers made up of 2 and 0 (e.g. 2002, 2020, 20002, 20200, etc.). eight out of ten of these numbers based on 2 and 0 were identified correctly, and overall 66% of the
numbers were identified correctly. Busnel does not draw any conclusions from the results other than whistlers can communicate even in difficult circumstances.

A more recent experiment was performed by Annie Rialland among the whistlers of Silbo Gomero (2005). She tested consonant identification by having a whistler perform a series of 68 nonsense vowel-consonant-tokens, which were then identified by another whistler 15 meters away. 15 different consonants and four different vowels were used in creating the tokens. The listener was able to correctly identify 57% of consonants and 87% of vowels. Rialland notes that there are broad categories of consonant types that are easily perceived, however. Coronal versus non-coronal consonants were categorized correctly 96% (62/64) of the time (i.e. one coronal or non-coronal consonant was confused for another). Also, voiced and unvoiced consonants were categorized correctly 87% of the time (56/64).

These categories (coronal/non-coronal and voiced/unvoiced) were first described in the work of Ramón Trujillo in 1978 and have since been used to teach new generations how to communicate using Silbo Gomero. By studying spectrograms, it was found that these two types of consonants are reliably indicated by the contour of the pitch at the transition between the vowel and the consonant. Coronal consonants are preceded by an upward pitch slide and non-coronal consonants are preceded by a downward pitch slide. Spectrograms indicate that voiced and unvoiced consonants are distinguished by a dip in the amplitude envelope or an interruption, respectively. By combining this knowledge based on spectral analysis and the results of the experiments, Rialland is able to bolster the argument that
listeners make use of these acoustic details to distinguish between sounds in whistled speech.

Consonants were the focus of Rialland’s experiment, and the vowels were not equally represented. No tokens with /e/ were used, /u/ only occurred 5 times, and /a/ was used far more frequently than any other vowel. Still, the results indicate the importance of testing theories of how speech surrogates are understood. Trujillo posited that there are two vowel levels in Silbo Gomero: acute (/i e/) and grave (/a o u/). But Rialland notes that there are consistent differences in F0 for the individual vowels across whistlers, even if there is occasional F0 overlap in production. In the experiment, the whistler did not merge the F0 of /a/ and /o/, and the listener was often able to differentiate between them, even though they are perceptually similar. Out of 85 presentations, /a/ was correctly identified 74 times and confused with /o/ four times, while out of 22 presentations /o/ was correctly identified 15 times and confused with /a/ five times. This indicates that not only are whistlers able to produce small differences between vowels consistently, but that listeners can perceive these differences.

The work of Busnel and Rialland demonstrate the inherent difficulty of studying musical speech surrogates both analytically and experimentally. Busnel had particular difficulty due to the disconnect between the experimental task and the cultural context of the speech surrogate. Experiments force people to listen and respond in ways that are unrelated to how they normally experience performances, so this work requires a deep knowledge of the culture and a sensitivity towards it.
Researchers are also confronted by the fact that the practice of many speech surrogates is diminishing. Riall and found it difficult to find whistlers who were good enough to replicate her experiment. Other authors have noted that speech surrogates are quickly dying out as their social function is replaced with new technologies (Meyer 2004: 410), which will make future studies more difficult to execute.

Difficulties aside, the study of musical speech surrogates has much to offer. Combining acoustic analysis and experimentation not only improves our understanding of how speech surrogates really work, but it can be used to create new pedagogical tools so that the practices live on. Musical speech surrogates also challenge our assumptions about the boundaries of music and language and provide a rich opportunity for the burgeoning field of music and language cognition. An fMRI study has shown that both the Broca and Wernicke areas associated with the production and understanding of language, were activated in the brains of trained listeners in Silbo Gomero while they listened to whistled sentences (Carreiras 2003). These same areas were less active in the brains of listeners who were not familiar with the whistle language. Such experiments are promising, but without more context the results tell us little about how listeners understand the verbal content of whistles.

Carreiras notes that whistle processing did show a difference from speech processing in whistlers. Specifically, there was less ventral-anterior temporal activation while listening to whistling compared to speech. This area has been
associated with analyzing and mapping sounds onto lexical representation. He suggests that this could be due to the smaller number of phonological contrasts in whistled speech, which requires less analysis. Whistlers showed similar activation in the posterior temporal cortex, which is associated with articulatory-gestural representations, during speech and whistle processing. Based on this, Carreiras concludes that “left-hemisphere temporal and parietal regions may provide complementary pathways for language processing” (31). How it is processed, remains unclear, however. During whistle processing, whistlers showed activation in premotor areas of the brain associated with the movement of tongue and lip. But do the whistles elicit the represented gestures of spoken words or the actions of whistling that then act upon lexical activation and selection? More research is necessary to better understand this interface.

1.4: The study of music and language

There has been increased interest in recent years about the relationship between music and language cognition. The most extensive review of this quickly growing body of research is Annirudh Patel’s book, *Music, language and the brain* (2008). Patel concludes, “As cognitive and neural systems, music and language are closely related,” and that “[c]omparing music and language provides a powerful way to study the mechanisms that the mind uses to make sense out of sound” (431). This is

---

7 These two processing streams, mapping sounds onto lexical representations and articulatory gestures, are discussed in detail in Chapter 4.
based on numerous overlapping processing systems shared by music and language, including: “the ability to form learned sound categories, to extract statistical regularities from rhythmic and melodic sequences, to integrate incoming elements into syntactic structures, and to extract nuanced emotional meanings from acoustic signals” (19). The research cited by Patel is based mainly on the relationship between music and spoken language, but only a small section of the book details musical speech surrogacy.

Patel mentions speech surrogacy in his discussion of pitch in music and language, specifically tonal languages. He notes that pitch spacing in speech is flexible, even in tone languages where the same tone might be realized on different pitches based on the speaker or the context of the utterance. For example, the baseline pitch and pitch range of an utterance trend downward over the course of a statement, yet listeners are still able to perceive the same lexical tones. Pitch in music is more stable and has a fixed spacing. Patel notes that despite this difference in the pitch spacing of speech and music, musical instruments can be used to communicate linguistic messages. His first example is the Lokele talking drum, which imitates the two tones of spoken Lokele with high and low tongues on a slit gong. The slit gongs do not have a fixed tuning scheme and different instruments have different intervals between the high and low pitches. To Patel, this suggest that “relations between linguistic tones are not conceived (or perceived) in terms of standardized intervals” (62). As another example, he cites whistle languages but says that it is unknown if they used fixed pitch intervals. Finally, he notes an Ethiopian speech surrogate
played on five-string lute, which mimics the fives tones of the language encoded. It is not known how important the fixed pitches are to the communication of verbal messages, but Patel suggest that if “standardized pitch intervals are not important for linguistic tones, one would predict that a good deal of mistuning would be tolerable before communication efficacy dropped below a predetermined threshold” (63).

Patel notes the research of D. Robert Ladd and others which suggests that while the pitch of lexical tones might change across contexts and speakers, the pitch level as a proportion of the current range remains relatively stable (58). In other words, the interval between lexical tones might change in terms of semitones or cents, but its relative position within a given range stays the same. This could explain why Lokele drum makers can make instruments with different sized intervals between the high and low pitches but still communicate messages. Flexibility in musical pitch might go beyond relative intervals within a given range, however. In my previous research, I found a great deal of flexibility in the pitch ranges utilized on the Hmong raj (Poss 2005). Not only do the intervals change between the pitches of a range, resulting in different scales, but as noted above the mapping of lexical tones onto the pitch levels within a range can change, too (e.g. the high-falling tone was the lowest pitch in one suab and the highest pitch in another suab). Despite such changes, verbal communication via speech surrogates remains robust, so communication must be happening through other channels. Other cues like rhythm and articulation and the performative context also provide a great deal of information.
The study of musical speech surrogates challenges our accepted notions about the characteristics of music because it requires research on musical systems outside of Western culture. Research in music cognition has long been plagued with a Western bias—rooted both the cultural backgrounds of the researchers as well as the ready availability of subjects for study (i.e. students at universities in the United States and Europe). This has begun to change as researchers have taken a more ethnomusicological perspective, but traces of it persist. For example, Ray Jackendoff takes a more pessimistic view of the relationship between music and language (2009). He notes that while music and language processing share some brain areas, “most of what they share does not indicate a particularly close relation that makes them distinct from other cognitive domains” (203).

While music and language both make use of pitch, Jackendoff argues that it is structured quite differently in each modality. For example, prosodic contours often rise and fall continuously over an utterance. By contrast, melodies have fixed pitches within a given space. Jackendoff characterizes this pitch space as “tonal” (199).

Elsewhere, Jackendoff and Fred Lerdahl have defined tonal pitch space as a system in which “every note of the music is heard in relation to a particular fixed pitch, the tonic or tonal center. ... Whether explicit or implicit, the tonic is felt as the focus of pitch stability in the piece, and melodies typically end on it” (2006:45). In tonal languages, “the tones form a fixed set that might be seen as analogous to tonal pitch space” (2009:200). Jackendoff notes, however, that the choice of tones is determined by the words, which means that no tone can function as a tonic or point
of maximum stability. Furthermore, tones drift within in an utterance, while musical
pitches are fixed in tonal space. This leads him to conclude that there are no
analogues in language to musical pitch space.

Research on musical speech surrogates challenges these points. Pitch in music and
language must be analogous to some extent for the systems to function. In Silbo
Gomero, the F2 of vowels corresponds to the F0 of the whistle (Rialland 2005). In
Hmong speech surrogacy, the lexical tones are consistently mapped to levels within
a given pitch range. In both cases, there is no tonal center to the pitch space. The
pitch contours of Silbo whistles are based on the vowels of the words
communicated. The melodies in traditional Hmong music, both vocal and
instrumental, are determined by the lexical tone of the word being performed. In
fact, lexical tone influences the melodies of music from a variety of cultures with
tonal languages (e.g. List 1961; Pugh-Kitingan 1984; Agawu 1998; Chapman 2001).

While the concept of tonal pitch space might be useful for describing the music from
some cultures, it is far from universal, and it reflects a bias based on Western
musical culture in which tonality is a well-developed theoretical concept. Jackendoff
attempts to draw on examples from other cultures to support his arguments, but the
ethnomusicological perspective is missing. Diverse musical practices can be found
that do not conform to the idea of tonal pitch space. But more importantly, many
cultures have their own ways of understanding the organization of musical pitches
without recourse to a tonal center. How are we to know if all people hear musical
pitches in relationship to one fixed pitch that provides stability? Jackendoff &
Lerdahl (2006) only cite research that is based on experiments using Western tonal music, so the question remains open.
Chapter 2: Hmong speech surrogacy

2.1: Location of the study

The Hmong people I worked with for this research are part of a relatively small community in the Upper Midwest of the United States, but they are inextricably linked to a diverse, global community that is connected by language, culture, and history. Knowing this context is necessary to appreciate the limits of my current research and the potential implications for future research.

The participants in this research and my previous research came from a belt of communities stretching from the Twin Cities of Minnesota to Madison, Wisconsin. According to the 2010 census, these two states were home to over 115,000 Hmong Americans. The number of Hmong people in St. Paul and Minneapolis was over 40,000, making it a leading center of Hmong American culture. California’s Central Valley is the other major population center for Hmong Americans. Out of the national total of 260,076, over 91,000 live in California with large concentrations in Sacramento and Fresno. Hmong Americans can be broadly grouped into two subdivisions: Hmong Der (Hmoob Dawb), also known as White Hmong, and Mong Leng (Moob Leeg), also known as Blue or Green Mong. These groups speak different dialects of the Hmong language, and while they are culturally very similar, there are some differences in terms of ritual practices. My initial contacts in the Hmong
American community were Hmong Der, and my language training was in the Hmong Der dialect. While I did not exclude Mong Leng people from my current research, all of the participants in the experiments were Hmong Der, with the exception of one. This is likely due to the method of recruitment, which relied on word-of-mouth from my previous contacts and the prevalence of Hmong Der people in the communities where the research was performed.

2.2: The history of Hmong-Americans

Hmong people began arriving in the United States after 1975. That year, the communist Pathet Lao took control of the government of Laos and began to persecute Hmong people who had fought against them for the past 15 years with the support of the American CIA. Several thousand were sent to re-education and labor camps and others simply disappeared (Duffy et al. 2004:8). Perhaps as many as 200,000 Hmong people escaped across the Mekong river to Thailand where they were placed in refugee camps while they awaited resettlement elsewhere (Lee, 2004). Over the next 20 years, before the Thai government closed the camps, Hmong people were relocated to several Western countries, including: Australia, Canada, and France. Around 119,000 were resettled in the United States where they eventually underwent a secondary migration to the current population centers to live near relatives and job opportunities (Bulk 1996). A final group of about 15,000 refugees arrived in 2003 when the Thai government shut down an unofficial camp
at a Buddhist temple. Around 460,000 Hmong people continue to live in Laos where they constitute around 7% of the population (Government of Laos 2005).

This resettlement pattern, based on waves of movement over the past 40 years, has resulted in a complex diasporic community. Hmong Americans include older people who grew up in Laos during the Secret War and others who were born in refugee camps where families sometimes waited for several years to be resettled. There are also a wide range of Hmong Americans who were born in the United States since 1975. Regardless of their age or culture, Hmong Americans are united by the history of the Secret War and the mass exodus from Laos. Increasingly, however, Hmong Americans are reconnecting not only to their relatives and coethnics in Southeast Asia, but also to the Hmong community in China where Hmong people likely originated.

There are an estimated 2.8 million speakers of various Hmong dialects in China where they constitute part of the Miao nationality (Lemoine 2005). Linguistic and genetic evidence suggests that Hmong people have lived in southern China for at least the past 2,000 to 2,500 years (Ratliﬀ 2004; Wen et al. 2005). Early histories of China mention a Miao tribe, which fought against the legendary Emperor Shun in the 3rd millennium BCE (Yih-Fu 1967), but it is not clear if Hmong people were a part of this. Well into the 17th century, imperial authors used the term Miao interchangeably with Man and Yi to refer to indigenous people on the southwestern frontier who would not submit to imperial rule, so it is impossible to draw ﬁrm conclusions. By the early Qing era, Miao people were being differentiated by their
cultural characteristics, like the color of women’s outfits, and depicted in “Miao albums” that included paintings and drawings based on accounts from provincial gazetteers (Diamond 1995:101-105). Although the terminology was more specific, it was still used inconsistently. By the mid-19th century, groups of Miao, in particular the White (Bai) and Green (Qing) Miao, can be associated with contemporary Hmong (Culas & Michaud 2004:67).

In the 18th century, Hmong people along with other ethnic minorities began to fight wars of resistance against the cultural and economic reforms imposed by the Qing dynasty (Mottin 1980:30-37). In search of independence, new farmland, and peace, Hmong people migrated from southern China to northern Vietnam around 1800 and subsequently moved to Laos, Thailand, and Myanmar (Burma) where they continue to live today. The Hmong that remained in China became part of the Miao nationality in 1949 during a systematic process of classification and identification of ethnic minorities. This was necessary to properly allocate seats for representation in government and to delimit areas of autonomous rule in the new People’s Republic of China (Schein 1985:74). The process was one of self-identification as well as classification by governmental ethnologists and linguists (Cheung 2004). Miao people were classified as one ethnic group with one language largely on the basis of linguistic data. It was also argued that the Miao people had a common origin and that superficial cultural or linguistic differences could be attributed to the separation of an original core population (Diamond 1995:108). The subgroups that make up the Miao nationality (such as the Hmong, Kho Xiong, Hmu, Genao, and A
Hmao) are not recognized at the national level, but many scholars continue to define them as separate ethnic groups since the languages they speak are not mutually intelligible and there is considerable cultural variety among them (Tapp 2002).

2.3: Previous research on Hmong music

Hmong Americans who grew up in Laos and Thailand brought with them musical practices that, at least to some extent, continue to be practiced in Southeast Asia today. Detailed information is lacking, however. Almost all of the research done on Hmong music has been done with Hmong people originally from Laos who now live in other countries, like Australia and the United States. Only Gretel Schwörer-Kohl has done extensive work among the Hmong living in northern Thailand where she studied the qeej, a bamboo free-reed mouth organ. The most complete picture of traditional Hmong music can be found in Amy Catlin’s several publications on her research with Hmong Americans. In the 1980s and 90s, she studied traditional vocal and instrumental music as well as new forms of Hmong popular music. But this represents only a fraction of the people who make up the Hmong diaspora. There have been no large scale studies comparing the practices of Hmong people who live in different countries and almost no research exists on the music of Hmong people in China.

Two works based on research from the 1920s and 30s provide the only view into Hmong music in China. One is a book of stories and songs published by David Crockett Graham who spent a great deal of time among the “Ch’uan Miao” in that era.
Graham was a missionary and anthropologist who lived among the Ch’uan Miao, who called themselves “Hmong” in their own language. Besides the song texts, his work tells us little about musical practice.

R. Gordon Agnew, an associate of Graham’s while he was teaching at the West China Union University in Chengdu. He worked as a dentist and accompanied Graham on some of his research trips. Based on this experience, Agnew wrote a detailed description of Ch’uan Miao music, including both instrumental and vocal music (1939). This publication includes numerous transcriptions of the melodies of songs. He even transcribed a melody blown on a leaf and a melody played on a guimbarde, as well as a handful of melodies played on the qeej. He provides us with what is likely the first published description of Hmong speech surrogacy: “The leaf is predominantly used in love-making. It may be played softly, or with it an individual may call loudly to a distant friend. For example [sic], the farmer working in the field may, with the leaf, call to his lady love, even when at a great distance. The leaf may also be used in times of trouble to summon members of the clan, since the sound, when played loudly, is strident and carries for a surprising distance” (15).

Agnew supplies little cultural context to the practices he discusses—focusing more on musical analysis utilizing the framework of Western tonal music, which is not well-suited to his task. But the genres he mentions and the situations for music

---

8 Agnew also describes performances on the guimbarde and qeej, but it does not appear that he understands that these performances communicate words. Even his description of leaf blowing suggests that he views it as a signal rather than a way to express a verbal message.
making align with traditional Hmong music in Southeast Asia, which suggests that these practices predate the migration out of China.

The paucity of Hmong music research makes it difficult to generalize across the Hmong diaspora. Since the music of Hmong Americans, and by extension the music of Hmong Lao and Thai, has been studied to the fullest extent, it tends to be presented as characteristic of Hmong people in general. This remains to be seen, although it is clear from the small amount of literature on Hmong music from other places that the basic instruments, song types, and musical genres are shared by most Hmong people who continue to practice traditional music.

2.4: Hmong musical instruments

The common traditional Hmong instruments include qeej (a free-reed mouth organ), raj (a family of flutes and free-reed pipes), ncas (guimbardes), nplooj (leaves), and spike fiddles that go by various names, including xim xaus. My previous research and this study are limited to the Hmong raj. Raj can refer to any type of tube but usually refers to those made from sections of bamboo. Some raj are used to carry water and others are used to hold small objects. Raj for music making constitute a diverse collection of aerophones commonly made from bamboo tubes, although other materials like metal and plastic can be used. Playing the raj is called tshuab raj, “tshuab” meaning “to blow.”

The raj nplaim has a brass free-reed that the performer blows through, giving the instrument its characteristic buzzy tone quality (Figure 2.1).
Raj nplaim can range in size from around a foot to several feet long, but usually have between five to seven finger holes. A small piece of bamboo or other grass can be used to make another free-reed pipe called raj lev les. The free-reed is cut directly into the body of the instrument and is covered by the performer’s mouth when blown. In Laos and Thailand, raj lev les were seen as disposable and they are not common in the United States.

The remaining raj for music making are end-blown flutes that go by a variety of names: raj hliav ncauj, raj pum liv, raj ntsha, and raj nploog. There may be a
connection between the morphology of the instruments and the names. In my previous research I found that raj hliav ncauj was usually used in reference to a fipple flute with five or more holes and raj nploog was used only in reference to a small flute with three holes. These terms were not used consistently between performers I met and when discussing recordings of flute performances, listeners used raj hliav ncauj, raj pum liv, and raj ntsha interchangeably. With more data, it is possible that more distinct categories would emerge, but clear distinctions between them do not appear to be of great importance to the musicians I have met.

All traditional Hmong musical instruments are used to communicate words, but the raj is the instrument better suited to study how pitch and articulation are utilized by listeners in understanding messages. For example, the most iconic of the Hmong instruments, a free-reed mouth organ called qeej, is mostly used for ritual performances—in particular, funerals and New Year festivals. Players memorize ritual texts and must study for years to master not only the words, but the proper way to sound them on the instrument. Untrained listeners cannot understand the encoded words, which, in the case of funerals, are directed at the soul of the deceased person rather than the mourners. Since the texts are memorized and vary little between performers, skilled listeners are more likely to rely on cues besides the acoustic relationships between the musical sounds and the sounds of the words to understand the message. In contrast, raj players usually reshape familiar expressions in new ways that address the context of the performance, which requires listeners to perceive small changes in the musical sound. Due to the limited
number of qeej players in the United States, it would also be very difficult to find sufficient subjects for a study.

The ncas, a guimbarde that is most commonly associated with courtship, is performed using the mouth as a resonator. This helps to preserve the vowel sounds and some of the consonants that are formed with the tongue. The similarity between the sound of the music and the sound of the language occurs on multiple levels: vowel, consonant, pitch, and possibly rhythm (although, this remains to be investigated). Since the sounds are very close to speech, listeners are likely to rely more the consonant and vowel information than the pitch information, although it may play a role, as well. It would be interesting to test the effect of pitch on understanding performances on the ncas by having listeners identify the words to phrases with manipulated fundamental frequency—perhaps raised or lowered for certain words or filtered out altogether. As is the case with whispered speech in tonal languages, the melody might not be as important as other cues, like vowel, consonant, and context in decoding performances (cf. Liu & Samuel 2004).

Nplooj, or leaves, are used as musical instruments. Performers hold a leaf between their lips and blow to create a shrill, high-pitched sound. Many Hmong Americans have stories about using nplooj in the jungles of Laos to signal while hunting or to communicate in code during battle. Leaves were also used to mimic natural sounds like the calls of birds or insects. Melodies can be blown on leaves and messages were communicated this way across large distances in the mountains. In the United States, leaf blowing is uncommon and usually only performed as a novelty or as a
cultural demonstration. While players use consistent intervals and scale types, exact control of pitch is difficult and melodies are irregular.

Hmong people also traditionally play a two-string spike fiddle. This instrument goes by a variety of names, including: *thaj chij*, *nkauj paj nruag*, and *nkauj puas ncas*, but it is most commonly known as *xim xaus*, which links it to the Lao and Thai versions of the instrument called *so* or *saw*. In my own research, I only encountered two musicians who played this instrument and only one of them had an instrument available to play. The melodies he performed were very similar in phrasing and scale-type to those he played on the *raj*. The *xim xaus* has not been well documented and while it is likely that it is used to communicate messages in ways similar to the *raj*, the low level of usage among the Hmong American community does not make it conducive to the current study.

Among Hmong Americans who grew up in Laos and Thailand, it is common to either play the *raj* or understand messages played on it. This is not the case among Hmong Americans who have grown up in the United States. I have yet to meet a young Hmong American who learned to play *raj* in this country, which raises the issue of whether or not the younger generation will take up the practice. Certainly, the older generation uses the *raj* for very different purposes than when they learned it in Southeast Asia where it was associated with courtship during adolescence. Many people I spoke with described becoming interested in the *raj* around the age they became interested in the opposite sex. Young boys, in particular, used the *raj* to say sweet things to young girls or to tease them. Things could be played on the *raj* that a
shy young boy would not want to say out loud. The instrument also provided some emotional distance if the object of affection did not return his feelings.

Young girls could also learn to play raj, but based on my conversations with older Hmong Americans, it seems that this was less common. Many women I met could understand words on the raj but not play it themselves. One woman described herself as a “naughty child” who disobeyed her parents and learned how to play raj herself. The process of learning described by most musicians was informal. Young people would listen to performances and ask other young people what was being said. Many people described being taught a few phrases by an older sibling or family member. It is clear, however, that the phrases commonly played on the raj are different from those sung during kwv txhiaj, a form of improvised, sung poetry associated with the courtship ritual of ball tossing, pov pob. The words played on the raj were not intended to be sung and were only vocalized during the learning process. They also do not have the typical rhyme pattern and repeated stanzas found in kwv txhiaj. In comparing the texts of kwv txhiaj and raj performances, there is greater flexibility and more repetition of shorter phrases on the raj.

Other circumstances for tshuab raj in Laos and Thailand were described by musicians. One person described playing a “morning song” on the raj as he walked through the village waking people up. He also talked about playing the raj as he walked through the jungle toward a new village. The sound would indicate his friendly intent. Many people remembered playing raj for entertainment or relaxation. And one musician describing competing in a raj competition in Ban Vinai
Refugee Camp, the largest of the many Thai camps that received Lao Hmong people after 1975. He won the competition by knowing more *suab* on the *raj* than the other players. *Suab* is often translated as “voice,” but in relation to the *raj* it refers to specific pitch ranges with associated melodic and verbal phrases.

In the United States, older Hmong Americans have no use for *raj* as an instrument of courtship, so many performances contain messages about the sadness of leaving Laos and Thailand and missing family who were left behind. Musicians now mainly play for their own enjoyment or to entertain other people, although it is also sometimes played in public for cultural displays. For example, it might be featured at a New Year Festival where a performer plays *raj* through a public address system for an audience. Not everything can be expressed on the *raj*. Performers limit themselves to a few topics, as noted above: courtship, the sorrow of moving from one country to another, and the sorrow of being alone—usually expressed in the metaphor of an orphan. Familiar phrases are reworked to address different audiences and are repeated and elaborated in new ways by each performer. The quality of a performance is assessed both in terms of the beauty of the instrumental sound as well as the clarity of the words expressed. While a basic pulse organizes most *nqes*, good performers play with some degree of flexibility—holding out notes

---

9 Ban Vinai was one of the largest refugee camps, with numbers approaching 45,000 people on just 400 acres of land. Many cultural practices of the Lao Hmong were continued in refugee camps, such as shamanism and foodways as well as musical practices (Long 1993).
for important words or to heighten the emotional effect of a phrase. Skilled players can also shift between different scales within the course of a single performance.

The raj can be seen as a middle ground in Hmong instrumental music. Like nplooj, xim xaus, and ncas, it is used in informal music making where messages are crafted to address the situation at hand. But like the qeej, it encodes the tonal information of the underlying utterance and a broad distinction between syllable onset types without information about the vowel. Agnew, in his 1939 article, noted that in Ch’uan Miao songs “the tonal progression—the basis of form—is fundamentally derived from the tone patterns originally fixed by the tonal structure of Ch’uan Miao speech” (21). But he did not work out the details of how the tones and the pitches were related. Eric Mareschal was the first to map this relationship in performances of kwv txhiaj (1976). Amy Catlin later verified the relationships he found between pitch levels within a given range and lexical tones (1982). My previous research built on their findings and found a mapping between lexical tones and musical pitch levels in tshuab raj nplaim that was very similar to their mapping. The most notable difference was finding that the falling breathy and falling creaky tone were performed on different pitches in predictable phonological/musical conditions (Poss 2005:141-2).
Table 2.1. The mapping of lexical tones onto pitches in *suab raj nplaim*. The highest pitch is labeled 1 and the lowest is labeled 4. Numbers in parentheses indicate alternate pitches utilized in certain phonological conditions.

<table>
<thead>
<tr>
<th>Lexical-tone</th>
<th>Pitch level</th>
</tr>
</thead>
<tbody>
<tr>
<td>high level</td>
<td>1</td>
</tr>
<tr>
<td>high rising</td>
<td>1</td>
</tr>
<tr>
<td>mid level</td>
<td>2</td>
</tr>
<tr>
<td>falling breathy</td>
<td>2 (3)</td>
</tr>
<tr>
<td>low level</td>
<td>3</td>
</tr>
<tr>
<td>falling creaky</td>
<td>4 (3)</td>
</tr>
<tr>
<td>high falling</td>
<td>4</td>
</tr>
</tbody>
</table>

Catlin actually applied her mapping of lexical tone and musical pitch in the transcription of a performance of *raj nplaim* (1982). She transcribed the melody, but she also wanted to know the words to find out if the mapping she and Mareschal found in *kwv txhiaj* would also be found in instrumental music. First, she sought help from her assistant, a young Hmong man from the same language group as the performer (*Moob Leeg*) who said he could understand the performance. According to Catlin, he analyzed the recording very closely, listening to it several times and writing down the words. Then she had the performer make his own transcription of
the words. He listened through three or four times and wrote down the words.

Finally, she had a relative of the performer make another transcription, but he was unsatisfied with his version, so he did it again with the help of the performer. Catlin noted that the number of syllables in the first two transcriptions did not match the number of pitches in the performance. Also, the tones of the words did not line up with the pitches in the expected lexical tone-to-pitch mappings. In contrast, the third version did match both the number of pitches and the expected mappings between the lexical tones of the words and the relative pitch levels.

Despite the different outcomes, Catlin says that both the listeners and performer thought that all three transcriptions were successful even though they did not use the same words. There was a general agreement of subject matter between all three versions and a good deal of overlap in terms of the phrases. In my own analysis of the transcriptions, I found that the final version consisted of 17 phrases and of these the first transcriber came close to matching 5. The other phrases the listener reported all elaborated the theme of having to leave one’s homeland and the sadness of leaving behind one’s family, one of the most common themes not only in *tshuab raj*, but other verbal arts, as well. The performer himself only transcribed 13 phrases in his version, all of which correspond to phrases in the final transcription except for one. Interestingly, they do not appear in the same order as the phrases in

10 It is not clear how the performer helped his relative perform the transcription. Catlin only says: "At first he was satisfied with his results, but eventually he decided to throw away his manuscript and begin again with the musician at his side. For several hours they labored with the tape, in the end producing text Version 3 whey [sic] they agreed was correct" (1982:188).
the final version. Again, the differences between the second and third transcriptions do not affect the overall meaning. Almost all of the phrases reported by the participants were also played or described by the performers I met during my field research. This demonstrates the limited nature of the raj repertoire since the performers I worked with were all from a different language and culture group (Hmoob Dawb) and were playing for me more than twenty years after Catlin made her recordings.

Although this type of transcription has little to do with how people actually experience performances in real life, it does bring to light different listening strategies. Listeners have a general idea of what to expect in the performance. The first transcriber likely recognized a few key phrases that are commonly performed by many musicians and then filled in the ambiguous phrases with ones that he was familiar with. It is impossible to know if he made use of lexical tone-to-pitch mappings or other features of the sound, but since the words of the transcription did not match the number of pitches in the performance it would seem that much of his interpretation was not based on these types of relationships. The performer's first transcription is much closer to the final version, but there are key differences. First, some of the phrases are not presented in the same order. Catlin is not specific, but her description suggests that he simply listened to the recording a handful of times, so it is possible he recognized the phrases but put them out of order while writing them down. This is indicative of the oral culture in which tshuab raj is performed. Repeated listening and analysis is not a part of the practice and it is not
surprising that performers would not produce an exact transcription. In my own research, I found that players were very good at explaining relatively short phrases or sets of phrases that they had just played, but that they struggled when asked to listen to a recording of an old performance and explain the words.

Besides being out of order, there were small differences in wording between the second and third versions. For example, the final transcription contains the phrase: “Tsis kho nej siab li peb yuas. Txiv leej tub tsi muaj niam muaj txi,” which translates to, “You (all) don’t feel lonely the same as us. I don’t have any mother or father.”11 Here the performer is talking about how his experience is different from that of Americans who have not undergone the loss of homeland and family. The references to not having a mother and father are related to the orphan trope that is found through Hmong song and folklore. In the second transcription, the performer reported: “Tsis kho nej siab es. Peb tsis muaj niam tsis muaj txi es.” In the first sentence, he drops the li peb yuas, which just means “the same as us.” In the second sentence, peb (we) stands in for txiv leej tub, a male, singular first-person pronoun only found in verbal arts like kwv txhiaj or tshuab raj. Peb and tub both are spoken with a high tone and performed on a high pitch in suab raj nplaim, the most scale type played on the instrument. The second version also has an extra tsis, the negative marker, before muaj txi (to have a father). This does not change the

11 I have standardized the spelling from the original text which was inconsistent and contained spelling errors. The phrase as it was documented by Catlin was: “Tsi kho nej siab li peb yuas. Tsiv leej tub tsi muaj niam muaj tsiv.”
meaning, but it likely does not correspond to a pitch in the actual performance. Small differences such as this could indicate that the performer is not really analyzing his playing note for note in the second version, but simply recalling the basic phrase once he recognizes the melodic outline. Exchanging pronouns, to take one example, is one of the common ways that performers adapt their messages to address specific contexts.

Two types of initial consonants are also encoded in *raj* performances. Hmong is a monosyllabic language and most syllables begin with a consonant or consonant cluster and end with a vowel. White Hmong has 53 consonants and 13 vowels. Annie Rialland analyzed a several recordings of *raj* performances that included a total of 86 different initial consonants (2005:263). She found that voiced nasals, laterals, and glides were performed as continuous sounds and that the other consonants, obstruents and voiceless sounds, were performed with an interrupted sound. I found a similar pattern in the performances I had recorded—strongly tongued articulations were used for words that began with obstruents, but transitions between sonorants were either lightly tongued on repeated notes or performed smoothly between pitches (Figure 2.2).
<table>
<thead>
<tr>
<th>Articulation</th>
<th>Consonant (RPA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tongued</td>
<td>c, k, kh, nc, nk, p, qh, s, t, ts, tx, z</td>
</tr>
<tr>
<td>Smooth</td>
<td>h, hl, hm, l, m, n, ny, v, y</td>
</tr>
</tbody>
</table>

Table 2.2. Consonants and their associated articulation types in *tshuab raj*. Consonants are listed in the Roman Popular Alphabet.

The exception to this pattern is /v/ (/v/). As an obstruent, it seems that it should be played with a tongued articulation, but both times it occurs in the recordings I have collected, it is performed with a smooth transition. Compared with other fricatives that have been found to be performed with a tongued articulation (/s/ and /z/), /v/ is the only one that does not use the tongue as an articulator, only the teeth and lips. Unfortunately, there is no documented performance of a word beginning with a similar phoneme, like /f/, to see if this holds true beyond this one case.

*Raj* performances consist of phrases called *nqes*, which are separated by a breath. Each *nqes* contains one or more complete sentences. Besides the words, players often insert formulaic melodic fragments that do not have a verbal meaning. This is especially common at the beginning of performances and at the end of each *nqes*, although extra pitches may be inserted within phrases to separate clauses or to add ornamentation. The following example is the first *nqes* from a performance a *raj nplaim*. It begins with a common melodic formula and ends with a closing melodic formula. The words, as explained by the performer, are: “*Lawv los lawv, yuav kho*”
*kho nej lub siab tsis kho?*” (Everyone, oh, everyone. Are you lonely or not?) Tongued articulations have been indicated with a “T” and correspond to the consonants /kʰʂ tʂ/. The other consonants, /l y n/, are marked by continuous transitions between the pitches.

![Figure 2.2. Transcription of a *nqes* performed in *suab raj nplaim* from a longer performance by Sai Pao Lee recorded in 2005, Minneapolis, Minnesota. Tongued articulations are indicated with a “T.” Rhythm and pitch are approximate.](image)

While it is clear that information about the lexical tones and onsets of the words intended by the performer are encoded in *tshuab raj*, it remains to be seen if these features are perceived by listeners and used to interpret words in the melodies. In

12 The lowest pitch in the range, played in this example for the high-falling tone as well as the melodic formulas, is often approached from a pitch around a quarter above. This optional pitch is never played in isolation but only as part of the falling melodic contour.
the experiments that follow, two different approaches are taken. In the first experiment, the effects of phrase length, pitch, and articulation on understanding the words of the performer are studied in a listening test. In the second experiment, the effect of pitch on the cognitive process of lexical selection is explored using auditory priming.
Chapter 3: Understanding phrases on the raj

3.1: Introduction

My previous study of the Hmong raj employed interviews, participant observation, and musical analysis to understand how performers communicate verbal messages on this musical instrument. Messages were found to show consistent relationships between lexical tones and pitch levels within a given range. I also found that the tongued articulation of pitches is associated with onsets that contain obstruent consonants. Performers limit their content to a handful of topics, like love and loneliness, which are associated with the typical contexts for performances (e.g. courtship or personal entertainment). Within these topics, performers utilize conventional formulaic expressions that can be modified to address specific circumstances, like mentioning the name of a potential girlfriend or using the singular or plural second person pronoun. Rhythm appears to be of lesser importance in communicating messages. For the most part, performers use a relatively steady pulse—sometimes lengthening important words and shortening less important ones. These rhythmic patterns were not necessarily consistent across performances. Sometimes a player repeats a message using a slightly different
rhythm without changing the meaning. Performances with a strict beat are characterized as unskilled and even humorously bad.

After my ethnographic research and music analysis the question remained: how do listeners make use of these elements to comprehend musical messages? Do they recognize the melodic contour of an entire phrase and infer a general meaning the way we might hear a familiar melody and think of the lyrics? Or do they analyze each constituent pitch and articulation (or groups thereof) for clues as to what the underlying message might be? I developed an experiment to investigate how the length of musical phrases affected the ability of listeners to understand the words intended by the performer. By analyzing the reported words that did not match those of the performer, I could also see if listeners used the pitch and articulation of the melodies to interpret verbal messages.

Rather than have listeners respond to entire performances, I based the experiment on individual phrases and shorter segments. I had encountered the use of short phrases during my previous field research. Musicians I interviewed would often play a short phrase and explain the meaning to me. Other interviewees explained that they had learned some phrases in a similar way—someone would play a phrase and explain the words. By using a more familiar form of presentation and shorter phrases, I hoped to encourage close listening to see if listeners were attentive to details of pitch and articulation. Still, I was concerned that the strangeness of the experimental situation might discourage subjects from responding, so I selected phrases that were common among the several performers I had previously met. I
used melodies performed by one musician and trimmed them to various lengths from 11 syllables down to 2 syllables. For the most part, one pitch on the raj is equivalent to one syllable, and since Hmong is a monosyllabic language, each syllable is one word.

Using typical experimental methodology, these items would be presented in a randomized order, but I decided to use a set order ranging overall, from long to short. The shortest phrases would be the most difficult to understand since they would be stripped of the context found in longer performances. I was concerned that if listeners were unsuccessful in the early part of the experiment, they would be more likely to stop or not provide a spoken interpretation of an ambiguous phrase. By presenting longer items first, I hoped subjects would experience success before attempting the most difficult items. As a result, the effect of the order of presentation cannot be tested.

While the possibilities for statistical analysis are limited, it might not be the best way to understand what is happening when listeners respond to phrases they hear. For example if subjects correctly identified 50% of the words in the items presented, it could not be inferred that listeners in general understand 50% of the words in actual performance contexts. Certainly, several strategies are used in listening to raj, and these strategies likely change depending on the context. A young girl being courted by a young man might be more attentive to the specifics of the message than someone listening for enjoyment at a New Year festival. Response accuracy could also be related to the length of the phrase. For example, in my previous field
research, I found that performers omitted words when explaining phrases on the *raj*. These words were not necessarily required to understand the message, but when speaking phrases they had just played on the *raj*, performers were more likely to omit words when explaining longer phrases than shorter phrases.

This could reflect different memory processes for music and language, but most likely it just reflects that performers are not greatly interested in word-for-word analysis of performances. The performers grew up in a culture that was primarily oral and analyzing performances after the fact was not a part of their experience. Some performers reported that when they were young, older siblings and friends would sometimes play a short phrase and explain what it meant, but breaking it down to the level of the word was not practiced.

The results were analyzed in several respects. The responses were checked to see if the subjects reported words that matched those given by the performer to see how effective people were in receiving the intended message. Words that did not match were analyzed to see how they related to the original message—were the words of the subject close in meaning to the original words or unrelated? These words were also checked to see if they matched the lexical tone-to-pitch mappings described above. This was done to find out if listeners were using the connection between pitch of spoken words and musical pitches to interpret the performances. The
responses were also checked to see if the onsets of the words given by the subjects matched the onset types indicated by the articulations in the recording.\textsuperscript{13}

3.2: Subjects

This experiment was performed in conjunction with the priming experiment described in the next chapter. My former Hmong language instructors in Madison, Wisconsin recruited subjects by word-of-mouth and I also drew on the Hmong Cultural Center and Center for Hmong Arts and Talent in St. Paul, Minnesota. Subjects were interviewed at the beginning of the experimental sessions. If they indicated they had experience playing a traditional musical instrument or they said they could understand messages played on musical instruments, they were offered the opportunity to participate in the comprehension test following the priming test.

Of the total 26 subjects who were tested, 13 played a traditional Hmong instrument or said they could understand musical messages. Two of these subjects began the test but stopped after a few items because they could not understand any of the musical samples played.

In my previous research, the Hmong Americans I encountered who played the raj or understood message played on the raj were all first generation immigrants who had been born in Laos and spent at least some of their childhood there. This pattern is also true of the subjects who participated in the comprehension test. Six were from

\textsuperscript{13} For a detailed discussion of the lexical tone-to-pitch mapping in \textit{suab raj nplaim} and onset-type see Chapter 2.4.1.
Xiangkhouang province, including two who were born in Long Cheng. The remainder was from adjoining provinces of Houaphan, Luang Prabang, and Vientiane, except for one subject who did not give a specific location and said he grew up in the Ban Vinai refugee camp in Thailand. Subjects ranged in age from 33 to 60 with an average age of 50. Only two of the subjects were female, which fits with my earlier observation that it was mainly young men who learned to play the *raj*. At the same time, I have encountered several women who know how to play and certainly many would have learned to understand the messages played by young men as part of courtship. The low number of women subjects could be related to recruitment rather than cultural practice, although it is difficult to pinpoint what would have caused the demographic slant. It is possible that since the recruiters for the older, musically experienced subjects were male, they might have been more likely to recruit male subjects.

In terms of musical experience, three of the subjects said they could understand musical messages but could not play any instruments. Most of the 10 remaining subjects played multiple instruments.

---

14 Long Cheng (*Looj Ceeb*) was a Lao-Hmong military base established by the American CIA during the Secret War. At its height in the early 1970s, Long Cheng was home to more than 30,000 Hmong people (Fadiman 1998:136).
Eight of the subjects indicated that they had experience singing. Only one said they could not play an instrument. All of the singers sang *kwv txhiaj*, except for one who only sang *nkauj*, or Hmong popular music that is based on Lao, Thai, and Chinese pop. In addition to *kwv txhiaj*, one singer sang *zaj tshoob* (wedding negotiation songs). Two subjects reported that they could understand *tshuab raj* but did not sing or play an instrument. These subjects performed very well in the experiment. Each one provided responses for 13 of the 16 items presented and match many of the words intended by the performer.

<table>
<thead>
<tr>
<th></th>
<th>Ncas</th>
<th>Qeej</th>
<th>Raj</th>
<th>Xim Xaus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 3.1. Instruments played by the 13 subjects who claimed to understand messages in traditional Hmong music.
These numbers support my previous observations on musical participation in the Hmong community. The ability to sing *kwv txhiaj* is common among by Hmong Americans who grew up in Laos and Thailand and there is a strong correlation between vocal and instrumental practice. Ritual genres like *zaj tshoob* and *txiv xaiv* are practiced by a few specialists. Also, the ability or inability to perform vocal

---

15 "Qeej nplog" translates to "Lao free-reed mouth organ." This refers to the *khene*, a free-reed mouth organ associated with Lao people. During my previous field research, I had met a few Hmong musicians who had learned to play this instrument. The few times I heard it played, the musicians used it to play Lao music, not Hmong music.
music is not directly related to the ability to understand messages played on musical instruments, further evidence that instrumental music is not based solely on vocal genres.

3.3: Experimental materials

Excerpts used for the experiment came from recordings of one performer, recorded in three different sessions using the same instrument. Due to the concern that the test would be too difficult for the subjects, since listening to small phrases out of context is not a typical part of Hmong musical practice, I chose commonly played phrases on the most commonly played instrument, *raj nplaim*, a free-reed pipe. In my previous research, I found that among people who played only one type of *raj*, it was *raj nplaim*. Many people played more than one type, but they all played *raj nplaim*. By using subsets of the available pitches on the instrument, *raj nplaim* can be played using a variety of *suab*, or pitch ranges associated with various topics and styles. I limited the experimental materials to the most commonly played *suab*. While it does not have an agreed upon name, many performers referred to it as *suab raj nplaim*. Within the limits imposed by the *suab*, performers can express a wide variety of ideas, so I limited the selection of phrases to those that I found to be played by several performers. More obscure phrases were tested in a follow up experiment described below.

A complete musical phrase on the *raj* is called *nqes* and might contain one or more complete sentences. I selected 12 *nqes* from recordings of a single performer and
extracted 16 excerpts ranging in length from 11 to 2 pitches. These excerpts consisted of commonly used expressions some of which were fragmentary.
<table>
<thead>
<tr>
<th>Item</th>
<th>Words</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lawv os lawv yuav kho kho koj lub siab tsis kho?</td>
<td>Everyone, oh everyone: Are you lonely or not?</td>
</tr>
<tr>
<td>2</td>
<td>Kuv tsis muaj niam tsis muaj txi.</td>
<td>I don't have a mother or father.</td>
</tr>
<tr>
<td>3</td>
<td>Nkauj Hmoob os nkauj Hmoob</td>
<td>Hmong girl, oh Hmong girl</td>
</tr>
<tr>
<td>4</td>
<td>Kuv tsis muaj kwv tsis muaj tig es.</td>
<td>I don't have any brothers.</td>
</tr>
<tr>
<td>5</td>
<td>Kuv tso kuv tus hluas nkauj.</td>
<td>I lost my girlfriend.</td>
</tr>
<tr>
<td>6</td>
<td>Koj yuav ciaj los yuav tuag.</td>
<td>You will become sick and die.</td>
</tr>
<tr>
<td>7</td>
<td>Kho kho kuv siab</td>
<td>I am lonely.</td>
</tr>
<tr>
<td>8</td>
<td>Leej niam leej txi</td>
<td>Mother and father</td>
</tr>
<tr>
<td>9</td>
<td>Muaj vi muaj ncaus</td>
<td>To have sisters</td>
</tr>
<tr>
<td>10</td>
<td>Koj puas paub?</td>
<td>Do you know?</td>
</tr>
<tr>
<td>11</td>
<td>Txiv leej tub</td>
<td>Young man</td>
</tr>
<tr>
<td>12</td>
<td>Nyob qhov twg es</td>
<td>To be somewhere</td>
</tr>
<tr>
<td>13</td>
<td>Paub tsis paub?</td>
<td>Do you know or not?</td>
</tr>
<tr>
<td>14</td>
<td>Zoo siab</td>
<td>To be happy</td>
</tr>
<tr>
<td>15</td>
<td>Pov tseg</td>
<td>To throw away (leave behind)</td>
</tr>
<tr>
<td>16</td>
<td>Koj hais</td>
<td>You say</td>
</tr>
</tbody>
</table>

Table 3.3. The words of the phrases used in the experiment, as described by the original performer, and an English translation

The excerpts were presented in the same order for each participant, starting with the longer phrases and ending on the shortest ones. Subjects listened to the excerpts through headphones and were asked to speak out loud the words that they
understood. Their responses were recorded for later review. Due to the difficulty of the task, subjects had the option to hear the excerpt more than once, but few requested it. Some items were more frequently requested for repetition than others (Table 3.4). There were no time constraints on responses.
Table 3.4. The number of requests for repetition of the experimental items.

<table>
<thead>
<tr>
<th>Item</th>
<th>Request for repetition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1</td>
</tr>
</tbody>
</table>

The responses of the subjects were recorded to computer via microphone and were transcribed later.
3.4: Results

One subject was not presented item 9 and another subject was not presented item 13. There was also one response to item three that was not decipherable during transcription. These individual incidents are excluded from the following analysis.
<table>
<thead>
<tr>
<th>Item</th>
<th>Words per item</th>
<th>% of subjects who responded</th>
<th>% of words reported</th>
<th>% of words matched</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>100 (11)</td>
<td>75.2 (91)</td>
<td>61.2 (74)</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>81.8 (9)</td>
<td>93.7 (59)</td>
<td>92.1 (58)</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>*90 (9)</td>
<td>97.8 (44)</td>
<td>82.2 (37)</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>81.8 (9)</td>
<td>73.6 (53)</td>
<td>69.4 (50)</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>81.8 (9)</td>
<td>87 (47)</td>
<td>72.2 (39)</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>36.4 (4)</td>
<td>70.8 (17)</td>
<td>45.8 (11)</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>100 (11)</td>
<td>100 (44)</td>
<td>97.7 (43)</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>72.7 (8)</td>
<td>96.9 (31)</td>
<td>68.8 (22)</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>*30 (3)</td>
<td>100 (12)</td>
<td>25 (3)</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>90.9 (10)</td>
<td>100 (30)</td>
<td>100 (30)</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>100 (11)</td>
<td>100 (33)</td>
<td>100 (33)</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>54.5 (6)</td>
<td>83.3 (20)</td>
<td>33.3 (8)</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>*70 (7)</td>
<td>100 (21)</td>
<td>50 (6)</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>54.5 (6)</td>
<td>**108.3 (13)</td>
<td>25 (3)</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>90.9 (10)</td>
<td>**140 (13)</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>45.5 (5)</td>
<td>**140 (14)</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>74</td>
<td>79.1 (125/158)</td>
<td>68.5 (549/802)</td>
<td>52.0 (417/802)</td>
</tr>
</tbody>
</table>

Table 3.5. Subject responses to the 16 items in order of presentation. Absolute numbers are in parentheses. % of words reported indicates the number of words reported by the subjects compared to the number of words given by the performer. % of words matched indicates the number of words reported by the subjects that matched the words given by the performer.

*Item 9 was not presented to one subject and item 13 was not presented to another subject. One of the responses to item 3 was not comprehensible. These are excluded from the analysis and affect the total number of words presented during the experiment.

**These values exceed 100 percent because subjects reported more words than were present in the original recording.
Despite the unusual task, subjects responded almost 80% of the time—meaning that they thought they could understand what the performer was communicating. The remaining 20% includes items for which subjects said they did either not understand the words or they thought the phrase did not have any words. Performers sometimes begin or end phrases with a few notes that do not have an associated verbal meaning, although they might be described with a syllable.

Subjects were presented 802 words total and reported 549 words (68.5%). There appears to be a correlation between the length of the item and the percentage of words reported. The average percentage of words reported for phrases with 4 or more words is 87.8% while the average for phrases with two or three is 108.1%. A percentage greater than 100 is possible because for these items, not only did subjects report the same number of words as the performer, but two of them reported more words than the performer originally gave.

The percentage of words that matched between the subjects and the performer also appears to be related to the length of the items, but not linearly (Figure 3.1). The three two-note items produced the lowest average percentage of words matched (8.3%). Out of the 47 words reported for these items, only three matched the words given by the performer. The results for the seven items with three or four words were more varied. Three of the three- to four-word items produced verbal responses that were identical or nearly identical to those of the performer. Another three of these items produced responses that matched only 32% of the words given by the performer (17/53). The last item in this group produced responses that fell
in the middle of the spectrum with 70.1% of the subjects' words matching those of the performer (22/31). The items with more than 4 words produced more consistent responses. Out of the 311 words subjects reported for items one through six, the average rate of matching the words of the performer was 83.0% with a range of 64.7% to 98%.

![Figure 3.1. The percentage of words reported by the subjects that matched those reported by the performer for each item.](image)

Of the 549 words reported by the subjects, 417 matched the words given by the performer (76.0%). This leaves only 132 words reported by the subjects that were different from those reported by the performer. These words were analyzed to see
if they matched the performance in terms of onset type and lexical tone-to-pitch mapping. This was done to find out if the subjects were using this information to figure out words that they did not recognize. 106 (80.3%) of the non-matching words matched the lexical-tone-to-pitch mapping of the words reported by the performer (Table 3.6).
<table>
<thead>
<tr>
<th>Item</th>
<th>% Matching T+O</th>
<th>% Matching T</th>
<th>% Matching O</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>52.9 (9)</td>
<td>52.9 (9)</td>
<td>52.9 (9)</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>85.7 (6)</td>
<td>100 (7)</td>
<td>85.7 (6)</td>
</tr>
<tr>
<td>4</td>
<td>66.7 (2)</td>
<td>66.7 (2)</td>
<td>66.7 (2)</td>
</tr>
<tr>
<td>5</td>
<td>50 (4)</td>
<td>62.5 (5)</td>
<td>62.5 (5)</td>
</tr>
<tr>
<td>6</td>
<td>83.3 (5)</td>
<td>66.7 (4)</td>
<td>33.3 (2)</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>100 (1)</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>22.2 (2)</td>
<td>88.9 (8)</td>
<td>33.3 (3)</td>
</tr>
<tr>
<td>9</td>
<td>44.4 (4)</td>
<td>55.5 (5)</td>
<td>77.8 (7)</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>41.67 (5)</td>
<td>100 (12)</td>
<td>41.7 (5)</td>
</tr>
<tr>
<td>13</td>
<td>13.3 (2)</td>
<td>93.3 (14)</td>
<td>13.3 (2)</td>
</tr>
<tr>
<td>14</td>
<td>90 (9)</td>
<td>90 (9)</td>
<td>90 (9)</td>
</tr>
<tr>
<td>15</td>
<td>95 (19)</td>
<td>100 (20)</td>
<td>95 (19)</td>
</tr>
<tr>
<td>16</td>
<td>57.1 (8)</td>
<td>57.1 (8)</td>
<td>64.3 (9)</td>
</tr>
<tr>
<td>Total</td>
<td>56.8% (75/132)</td>
<td>78.8% (104/132)</td>
<td>59.1% (78/132)</td>
</tr>
</tbody>
</table>

Table 3.6. Words reported by the subjects that did not match those reported by the performer that did match in tone and onset, just tone, or just onset.

Matching between the onset type of the response and original was less frequent. Out of the 132 responses that did not match the words of the performer, only 81
matched in terms of onset type (61.4%). My previous research has found two types of articulation that are associated with different onset types. This suggests that while the articulation of pitches can be useful in resolving ambiguous messages, the pitch level is more salient. Articulation, tongued or non-tongued, only provides broad distinctions between onset types, but the pitch level provides more information. The importance of pitch content is also apparent in interviews with performers who are keenly aware that the sound (*suab*) of the word must match the sound (*suab*) of the *raj*. In contrast, none of the performers I talked with were aware of their use of consistent articulation patterns. When resolving ambiguous phrases, directing attention toward the pitch level seems to be the most important strategy.

3.5: Discussion

The results suggest that skilled listeners are highly successful at interpreting musical messages even when they are taken out of context. Shorter phrases make identification of the performer’s underlying words more difficult, but listeners use different strategies to extract verbal meaning, even if that meaning does not match the exact words intended by the performer. Subjects reported a smaller percentage of the words for longer phrases than for shorter phrases. The higher number of words in the longer phrases increases the probability that subjects will not provide a response for each one. But this same pattern was found during my previous field research when players would perform a phrase and then explain the meaning. Often, they left out words that did not affect the meaning or that were repetitious. This was
also true for the subjects in the experiment, even when they reported nearly identical phrases as the performer. For example, in response to item one (“Lawv os lawv yuav kho kho koj lub siab tsi s kho”) one subject reported “Lawv los lawv kho koj siab tsi s kho”—omitting yuav, the repetition of kho and the classifier lub. None of these changes affect the meaning. It is simply more succinct.

The difference between the number of performed words and those reported by the performer could be related to the differences between producing and understanding speech and speech surrogates. The repetitions that might help to disambiguate similar surrogate phrases are unnecessary in regular speech because the words are clear. It could also be that performers develop a repertoire of favorite phrases over time and do not think of their phrases one word at a time in performance. Similarly, listeners do not need to understand messages word-for-word for the communication to be successful, as noted in Catlin’s original work. It is likely that this type of listening and understanding is typical of actual performances of tshuab raj, which are made up of several long phrases that repeat and vary the same ideas within a given topic.

While testing more subjects could better establish the pattern of responses to phrases of different length, as well as equalizing the number of items for the different phrase lengths, it may not be possible to discern the source of variation. This also points to a problem inherent in this type of experiment: the goal of listeners may not be to understand every word in a performance. Also, listening to phrases out of context and then reporting the words is outside of their normal
experience. The results of this analysis are limited to noting that it is possible for some listeners to listen analytically to long phrases and that, overall, listeners are successful in understanding the general message of the performer.

The percentage of words reported by subjects that matched those of the performer varied by length of the items. The items with two words produced few responses that matched those of the performer while items with 6 or more words produced a more consistent percentage of words that matched (83.0%). Items with three or four words produced responses that matched very well or matched very poorly, with one item that fell in between. In my previous field research I noted that performers usually taught me phrases at the three- to four-word level. Two-word expressions, like “nkauj Hmoob” (“Hmong girl”) or “niam txiv” (“parents”) were invariably expanded through repetition (e.g. “nkauj Hmoob os nkauj Hmoob”) or stated in parallel expressions (e.g. “leej niam leej txi”). More elaborate poetic language is also employed. “Girl,” “nkauj,” becomes “tus hluas nkauj” and “boy,” “tub,” becomes “txiv leej tub,” a more poetic way of saying the same thing. This verbosity reduces homophony. Instead of one high pitch for the high-tone word “tub,” a distinctive three-note melody is heard leaping from the highest note to the lowest note and back up to the highest note for “txiv leej tub.” These expressions are also repeated several times during a performance, which aids the listener in recognizing them. In my previous research, I stated that it was unlikely that people could reliably identify one or two note excerpts from raj performances since the context would be insufficient for ruling out homophones. This seems to be borne out.
in the results of this experiment. Listeners were unsuccessful in getting the words intended by the performer for the two-word items.

Compared to two-note phrases, three- to four-note phrases produced more inconsistent responses. Responses were evenly divided in nearly matching all of the performer’s words or very few, with the responses to one item falling in between. At this level, it is clear that listeners are able to recognize familiar phrases and can successful interpret the original message. If they come up with a phrase that is different from the performer, however, there is not enough context to redirect them to the phrase intended by the performer. This results in some items with very low matching scores. The responses to item eight fall in between these extremes. Only one subject out of the eight that responded reported words unrelated to those the performer had given. Of the remaining seven, four matched the words exactly (“leej niam leej txi”) while the other three replaced “leej” with “koj.” This changes the meaning of the phrase from “parents” to “your parents,” but both “leej” and “koj” are performed on the same pitch level. They would be articulated differently—“koj” would be tongued while “leej” would not—but there is no other context to lead the listener to prefer one version over the other.

Responses to phrases longer than four words were variable within a limited range. The percentage of words matched was not particularly low, but it was never 100%. As discussed above, this could be related to dropping unnecessary words when reporting long phrases. For example, five of the subjects who responded to item two left out the repetition of the negative particle “tsis” from the original phrase “kuv tsis”
muaj niam tsis muaj txiv." This does not change the meaning of the phrase ("I don’t have any parents.") Listeners also might not be able to pick out or recall all of the words as phrases become longer. Even if they miss a few words, there is more context, which can help to disambiguate homophonous expressions. Since there are more words, there is also a higher probability that they will recognize a melodic outline and be able to infer a familiar phrase. For the most part, the listeners reported different words from the performer in situations where ambiguous homophony could not be resolved.

For the shortest items, those with two or three words, subjects sometimes reported more words than would be indicated by the melody. There are several possible explanations for this discrepancy. In longer phrases, there is more opportunity for subjects to omit words, some of which may not be necessary to convey the meaning of the phrase. For example, the first item communicates the following words: Lawv os lawv yuav kho kho koj lub siab tsis kho. Three subjects left off the first three words, which are often used as an opening formula as a call for people to listen. Other subjects left out words that do not affect the meaning of the phrase, like the repetition of kho or the classifier lub.

In contrast, some subjects reported more words than pitches heard for the two-word items. For item 16, “koj hais,” one subject reported “kuv tsis muaj niam muaj txi” ("I don’t have any parents"), a difference of four words. In this case, “koj hais” and “muaj niam” would be played on the same pitches, which might have led the subject to think that it was part of the parallel expression “muaj niam muaj txi” ("to
have a mother and father” or “to have parents”). In tshuab raj, performers often present themselves as an orphan to emphasize their loneliness, so it is more common for performers to say that they do not have parents than to say they have them. It is not surprising that the subject in this case put the phrase in the more common negative form: “I don’t have any parents.” This same subject gave the same response, “kuv tsis muaj niam muaj txiv,” earlier in the experiment to item four, “kuv tsis muaj kwv tsis muaj tig es.” The phrase could have been primed for recall when the fragment with a related melodic contour was presented. Even if this explanation does not explain this subject’s response, it is worth noting that the ambiguous two-note phrase brought to mind a complete phrase, rather than fragment.

In the other case where a subject reported more words than presented, there was a glitch in the playback of the audio file that made the two notes of item sound like three. The performer’s intended content for this excerpt was “zoo siab” (“too be happy”), a sentiment that is less common in tshuab raj than “kho siab” (“to feel lonely”). “Zoo siab” and “kho siab” both have the same pitch and articulation pattern in suab raj nplaim and 4 of the 6 subjects who replied to this item reported the word “kho,” while none reported “zoo.” The subject who reported the extra word responded with “kho kuv siab” (“I am lonely”), which matched the melody and articulation pattern in the playback of the file.

Subjects were highly successful in selecting words that matched the lexical tone-to-pitch mapping when their words did not match those of the performer. Considering that there are four possible pitch levels, the result is much higher than pure chance,
which suggests that listeners were using the sounds of individual pitches to interpret the words. For item 12, “nyob qhov twg es” (“to be anywhere”), one subject reported, “kuv yuav tso es” (“I will leave”), a phrase totally unrelated to the content of the reported original but matching in terms of how the lexical tones would be realized melodically. The high accuracy in responding with matching tonal outline demonstrates that listeners are highly sensitive to the pitch content of musical phrases.

In some cases, these words were unrelated to what the performer had originally said, but subjects often replied with words that had a similar meaning or function. For example, “lawv os lawv” (“everyone, oh, everyone”) in item 1 was matched by four subjects while four other subjects instead reported “Hmoob os Hmoob” (“Hmong people, oh, Hmong people”). Both expressions are intended as calls for people to listen. The phrases also match in how they are played on the instrument. “Lawv” and “Hmoob” are both played on the highest pitch in suab raj nplaim and neither is articulated with the tongue since they begin with sonorants. Since performances can begin with either saying, and they are both played the same way, there is nothing about the context that would favor one interpretation over the other. In the fourth item, “kuv tsis muaj kwv tsis muaj tig es” (“I don't have any relatives”) one subject substituted the first-person plural pronoun “peb” (“we”) for
the first-person singular pronoun “kuv” (“I”). As in the previous example, both would be performed on the same pitch and with the same articulation. The context that would indicate the number of the pronoun is not available since the phrase is out of context. Several subjects exchanged the “tsō” (“to lose”) of the fifth item “kuv tso kuv tus hluas nkauj” (“I lost my girlfriend”) with “nco” (“to miss”). These two words have similar connotations as well as being performed the same way in terms of pitch and articulation.

It appears that listeners sometimes did not base their responses on lexical tone-to-pitch level mapping, but rather on complete phrases. This is evident in responses to item nine, “muaj vi muaj ncaus” (“to have sisters”). Only three subjects reported any words and none matched the words of the performer exactly. One subject reported “leej vi leej ncaus,” a parallel expression that simply means “sisters.” The melodic outline and articulation is the same on the raj for both phrases, so there is nothing to suggest “muaj” over “leej.” Another subject reported “leej kwv leej tig” (“brothers/relatives”) which has would be played on the raj with a similar contour, but kwv would be one pitch level up from vi and tig would be one pitch level up from ncaus. “Leej kwv leej tig” is more common than “leej vi leej ncaus” or “muaj vi muaj ncaus” because it can refer to all of one’s relatives not simply brothers. It is possible that the subject misheard the recording or was reminded of the more familiar phrase and reported it without noticing the difference in the final pitch. The other

16 Of the 9 subjects who responded, 3 reported “kuv,” the word given by the performer, and 5 omitted the opening pronoun altogether.
subject reported “tsis muaj niam txiv” (“I don’t have any parents”), which does not fit the melodic outline or pattern of articulations found in the recording. “Muaj niam” and “muaj ncaus” would be played to the same pitches, which might have prompted the response of the subject. Also, the item immediately prior to this one was “leej niam leej txi” (“parents”) to which the subject responded “koj niam koj txi” (“your parents”). This could have brought to mind the usual context for mentioning parents in tshuab raj—not having them.

While listeners are sensitive to pitch, they are comparatively insensitive to onset type. This is not particularly surprising: the articulation of individual pitches only distinguishes between two broad categories, for the most part: obstruents and sonorants (see Table 2.2). Since White Hmong has 53 consonants, this bipartite division provides little help in narrowing down possible words. If it is not perceptually important, why is it performed reliably by several different performers? One possibility is that the performers automatically move their tongue to stop the airflow as they imagine the words that they play and that this technique has simply become part of the style that is unconsciously passed from performer to performer. As noted in Chapter 2, words that begin with /v/ are not played on the raj with a tongued articulation even though it is an obstruent. Since it does not use the tongue as an articulator when spoken, perhaps there is not an impetus to use the tongue when playing it on the raj. However, /p/ is also an obstruent that is articulated with the lips, yet it is articulated with the tongue when played on the raj.
Does its status as a stop consonant encourage the performer to stop the sound of the instrument with the tongue?

Rather than focusing on the method of articulation, Rialland highlights the shape of the signal envelope: continuous versus interrupted (2005:263). She found obstruents and voiceless consonants to be realized on the raj with a period of silence, while voiced nasals, laterals, and glides were performed as continuous sounds. Rialland contrasts this with Silbo Gomero, which does feature periods of silence even in the realization of continuous consonants. She goes on to say that the sharp contrast found between the two onset types in tshuab raj “might be related to the fact that the coding of consonants by a more subtly varying array of signal-envelope shapes in a tone-based whistled language might interfere with the recognition of contour tones, which require precisely marked beginnings and endings.” In other words, a wider range of articulation types might make it more difficult for listeners to pick out the pitch, which is the more important part of the signal. But it could also be that there are only two types of onsets because the encoded lexical tone, in the context of a familiar set of topics and phrases, is sufficient for listeners to understand the intended message. The extra information might not be necessary for people to achieve effective communication. Tongued articulations might be an artifact of how the tongue is used to generate obstruent consonants or simply a part of the style that is passed along without a clear origin.

Beginning with the earliest accounts of speech surrogates in scholarly literature, the similarities between speech and musical have been a primary focus of research. The
implication has been that listeners perceive these similarities and use them to interpret the underlying words. At the very least, these features are seen to stand in as a symbol of the message. In *tshuab raj*, it would appear that similarity in articulation between the spoken words and the musical signal is not used by listeners, at least to a great extent, to recognize what the performer is saying. An interesting follow-up would be to create synthesized versions of the experimental materials and see how it changed the pattern of responses. It seems likely that the results would not be very different since the similar *suab* and collections of sayings are performed on the two-string fiddle, *xim xaus*, as well as leaves played as free reeds (*nploof*). For future research on other speech surrogates, the results indicate that more attention needs to be placed on what is salient in the signal for listeners, not just on what features speech and music have in common.

Pitches are more important for understanding *tshuab raj* than articulations, but are these pitches bringing to mind individual words or does the sequence of pitches bring to mind longer phrases that incorporate the fragment? Or is there interplay between these approaches? The results do not provide a clear answer, but there is evidence that some subjects responded in terms of phrases rather than individual words. When subjects did not report words that matched those of the performer or that matched the melodic outline, they still responded with phrases or fragments of phrases. For item 8, “*leej niam leej txi,*” one subject reported “*koj puas paub?*” (“do you know?”). This is a very common phrase and was actually used for item 10, for which the same subject also reported “*koj puas paub?*” “*Koj puas paub?*” matches the
melodic outline of "leej niam leej txi" without the second leej. It seems unlikely that the subject would not have noticed the third pitch in the phrase, which is the lowest on the raj nplaim and has characteristically buzzy timbre. Yet his or her response ignores this note for the sake of keeping the phrase intact. As noted above, another subject reported an entire six-word phrase in response to a two-note excerpt. The integrity of the phrase was maintained even though it did not relate to the length of the stimulus.

Item 15 was the serial verb "pov tseg" ("to leave behind"), which is usually used as part of a longer phrase like: "Kuv tso kuv leej niam leej txiv pov tseg" ("I left behind my parents.") It always occurs at the end of the phrase. None of the 10 subjects who responded reported these same words, but all of them reported two words that matched the melodic outline. They also all began their responses with the first-person singular pronoun "kuv" as part of a fragment that would be used as the beginning of a statement. One reported "kuv hais" ("I say") while three reported "kuv tsis" ("I don’t"), which was used earlier in the series of items. Six reported "kuv tus" ("my (classifier)"), which is most often used as part of the noun phrase "kuv tus hluas nkauj" ("my girlfriend.") This was used in item five, "kuv tso kuv tus hluas nkauj" ("I lost my girlfriend") and all of the subjects who reported "kuv tus" had reported it as part of their responses to that earlier item, which in general matched the words of the performer. Whether they had this phrase in mind or not, it seems likely that they had a longer phrase in mind, rather than simply two words, like the serial noun that was part of the performer’s original version.
Again, the results of the experiment cannot prove that listeners think in terms of phrases or individual words. Future experiments could explore this idea further by offering subjects very short melodies and asking them to complete the phrase. This could help to establish the length of phrases that listeners have in mind when they listen to *tshuab raj*. It would also help to better establish which phrases are most commonly expected. Since the subjects would be using their imagination to complete the line, their responses might also help to determine if listeners have in mind verbose versions of phrases (*i.e.* versions that include the repetition and elaboration that is common in performances, but less common in reports of what words are present in longer phrases.) The phrases used for this experiment, however, represent only a handful of well-known phrases that are utilized in traditional *raj* performances. Expert performers can play elaborate, sometimes virtuosic melodies, often switching between different styles. How are these performances interpreted by listeners? How do they relate to common phrases like the ones used in the experiment just described?

Part way through the field experiments, it became clear that listeners did not find the experimental setting an impediment to understanding musical phrases. I was interested to see how subjects would approach longer or more ambiguous phrases (*e.g.* phrases that began in the middle of a sentence), including phrases played on *raj*.

---

17 Alison Wray, among others, has argued that the mental lexicon contains not only small, combinatorial units, but also longer formulaic phrases (Wray 2008). It is possible that phrases played on the *raj* are stored and retrieved as units like other forms of formulaic language. The implications for future research are discussed in Chapter 5.
different instruments and in different *suab*. I created a second test using phrases of this type, but I was only able to administer it to two subjects. Their responses followed a similar pattern: long phrases were described with a minimum of words, catching the main meaning; ambiguous phrases were not understood; phrases that contained familiar expressions were understood even when played on different instruments in different *suab*. This is purely anecdotal and more testing will be necessary to see if this pattern holds for more listeners.
Chapter 4: Priming the mental lexicon with pitch contours

4.1: Introduction

Are relationships of similarity between speech sounds and the sounds of speech surrogates utilized in understanding Hmong musical messages? Or are these features vestiges of past practices now purely symbolic? The comments of musicians would suggest they are aware of the relationship between the sound of words and musical sounds. Is there an automatic, perceived similarity between the pitch of instrumental sounds and the lexical tones of words? Or are sounds matched in the sense of a set of consistent relationships between lexical tones and musical pitches that were learned through enculturation?

In the Hmong language, the word suab is used to refer to lexical tones. More generally, suab can mean voice as in the human voice or the sound of a musical instrument. It can also refer to a particular style of song or instrumental performance defined by conventional pitch collections and musical conventions (e.g., opening melodic formulas, song topics, etc.). But it can also refer to the noises of animals and sounds made by things like guns. Hmong musicians talk about matching the suab of words with the suab of the raj, but his does not necessarily mean that the pitch is the same. At least two of the lexical tones in Hmong are marked by other features than pitch. Voice quality may be more salient for
identifying the low-falling breathy/creaky tone as well as the low-falling tone ending with a glottal stop (Ratliff 1992). These vocal qualities are not preserved in musical realizations, nor are the pitch contours of the tones, except in one case. The high-falling tone is often realized as a two-pitch progression moving from high to low between pitches about 50 cents apart when played on the raj. In performance, musicians often ornament individual pitches with small upward glides, but these glides are not correlated with the tones of the underlying words (Poss 2005).

Still, Hmong speech surrogates are at least in part based on relationships of similarity between the pitches of musical and speech sounds. These similarities might give listeners an advantage in discerning words in musical performances, either by activating or inhibiting the activation of candidate words in the mental lexicon. Research on the role of pitch in the perceptual processing of lexical tones and in spoken and visual word recognition has increased in recent years, but it remains uncertain how the tonal information in speech is mapped onto mental lexicon and how it constrains lexical activation. In spite of this uncertainty, it is clear that prosody, and lexical tones in particular, are crucial in language processing—holding out the possibility that pitch, as realized on musical instruments, might affect lexical access and selection, helping listeners to understand the encoded words.

How is tone used to access the words in the mental lexicon? While there are several current models of speech recognition, including TRACE (McClelland & Elman 1986), Cohort (Marslen-Wilson 1987), Shortlist (Norris 1994), and the Neighborhood
Activation model (Luce & Pisoni 1998), most are based on the same basic process: activation of lexical items based on the acoustic input and competition between those items for selection. The models differ in how similarity between the input and the items is mapped and what items are activated, but they mainly focus on segmental phonology. Only Shortlist takes prosody into account as a cue for word segmentation, but it only considers stress and duration and does not include lexical tone.

While they are not well-represented in speech recognition models, lexical tones are lexically contrastive and must play a role in lexical processing for speakers of tone languages. Changing the tone of a syllable can change the meaning of a word. For example, in Mandarin Chinese, the syllable *ma* can mean ‘mother,’ ‘hemp,’ ‘horse,’ or ‘to scorn’ depending on whether it is pronounced with a high-level tone, mid-rising tone, low-dipping tone, or high-falling tone. A variety of studies indicate that tone is implicated in lexical processing even though it remains unclear how tone is integrated into the mental lexicon. One of the first such studies was a categorization experiment by Fox and Unkefer (1985). Subjects identified two tones in Mandarin Chinese along a continuum using tokens that were either words or pseudowords. The results showed that Chinese subjects stretched the boundaries of the tone categories if the change from one tone to another changed the token from a word to a pseudoword compared to a continuum where the shift was between two words. English speaking subjects did not show this effect, suggesting that the result was
related to lexical processing. The study did not address the question of how tone is processed—whether it is used to activate lexical candidates or select among them.

Cutler and Chen (1995) found that a mismatch in tone or rhyme had similar priming effects on word recognition in Cantonese. The materials consisted of bisyllabic words and pseudowords which were present in prime-target pairs where there was a match or mismatch of tone or rhyme in the first or second syllable. Responses were slower when the second syllable mismatched in either tone or rhyme compared to responses following an unrelated prime. In contrast, responses were faster than the control when the mismatch was in the first syllable. Inhibition was interpreted as evidence of competition between simultaneously activated words, i.e. the match of tone or rhyme in the second syllable was sufficient to activate candidate words. They attributed facilitation when the mismatch was in the first syllable to the form overlap between the second syllables of prime and target.

Other studies suggest that stress and pitch accent are sufficient to constrain lexical activation. In a 1999 study, Cutler and Otake showed that Japanese words that differed only in pitch accent did not cause priming in a lexical decision task. The difference in pitch accent was seen as limiting the initial activation of the accent pair. A similar study in Dutch reached the same conclusion. Cutler and van Donselaar (2001), using a lexical decision task, found that segmentally identical words differing in stress did not prime each other. Their conclusion was that the suprasegmental information was able to resolve segmental ambiguity for the subjects.
Conflicting results have been found in shadowing studies of Cantonese. Yip, Leung, and Chen (1998) found that repeating monosyllabic target words was facilitated when they differed only in tone (i.e. the prime and target matched in onset and rhyme). They interpreted this to mean that tone was not sufficient to limit the activation of words that were segmentally identical but different in tone. Yip (2001) repeated the experiment using more filler items to reduce the number of phonologically related trials, which could make subjects more ready to produce a spoken word (Goldinger et al. 1992). He found the same facilitation for words that matched in onset and rhyme, but also found facilitation for words that matched in rhyme and tone. Yip attributes this advantage to the rhyme and not the tone, since Slowiaczek et al. (2000) found a facilitation for rhymed words in a shadowing task. He goes on to argue that the results suggest speakers of Cantonese are more sensitive than Mandarin speakers to segmental information and that the syllable cannot be decomposed in the processing stage.

A more recent lexical decision experiment in Mandarin found mixed results. Lee (2007) tested primes and targets that were either identical, unrelated, or that shared the same segments or tone. He used both direct form priming and mediated priming in which the prime was related to the target by a third word that was not presented (e.g. lou3 - jian4zhu0 'building', where lou3 is related to lou2 'hall', which is related to the target semantically.) In the direct priming paradigm, segmentally identical primes and targets that differed in tone did not produce facilitation, in line with previous work by Cutler and other, regardless of whether the stimulus onset
asynchrony (SOA) was 250 or 50 ms. In the mediated priming paradigm, facilitation was found for minimal tone pairs at the 50ms SOA. This is same condition in which Yip and others found facilitation in Cantonese speakers, but he used an SOA of 250ms and a shadowing task instead of lexical decision. Lee attributes the facilitation at the shorter SOA to an early activation of the minimal tone pairs due to form overlap. Somewhere between 50 and 250 ms, tonal information is used to rule out incompatible candidates, eliminating the facilitatory effect. There was no facilitation using the 50 ms SOA in the direct priming paradigm. Lee suggest that this is because the prime is a strong competitor against the target based on form similarity, which could slow down the lexical decision.

In a subsequent shadowing experiment using Mandarin monosyllables, Poss et al. (2008) found that primes that matched target words only in lexical tone (i.e. not in segments) inhibited spoken responses compared to targets and primes that did not match segmentally or in lexical tone. This result was replicated in a lexical decision task using the same experimental materials in which subjects pressed a button to indicate whether the target was a word or not. In both experiments, the inhibition was found for words only, not pseudowords, indicating that the effect is related to lexical processing. These results suggest that lexical tone might have more of a role in lexical selection than inhibiting the activation of candidates that match segmentally but differ in tone. Poss et al. concluded that the effect might be caused by a group activation of words that are related by tone—increasing competition and delaying lexical selection.
These studies raise important questions that remain to be answered. Are the differences between the experiments due to differences in language? Mandarin has four tones while Cantonese has six. It is possible that tone is less important in Cantonese since many of the tones are perceptually similar. This could have implications for the Hmong language, which has eight lexical tones. Also, how does the perceptual similarity of tones affect priming? Lee notes that other studies have shown that the two most acoustically similar tones in Mandarin, the rising and low falling-rising tones, are more easily confused than the other tones (e.g. Gandour 1983).

The results of the above experiments are based on experiments that use syllables as primes. Speech surrogates, like those played on the raj, are based on musical sounds that preserve only a few features of speech. It is not clear how the processing of speech surrogates relates to the processing of lexical tone in speech itself. The following experiment investigates this question with a shadowing task in Hmong using pitch contours as primes for word and pseudoword targets.

4.2: Materials

Primes were created based on four tones from White Hmong, which are performed on the raj: high-level, low-level, high-falling, and mid-rising. These tones were selected because they are differentiated mainly by pitch, as opposed to the falling-breathy tone, which has a particular timbre, or low-falling with a glottal stop, which is often short in duration. These tones are also contrasting: level versus moving, and
high versus low. Two sets of primes were generated based on these four tones: vocal and instrumental. One set consisted of pitch contours taken from words spoken by a native male speaker, who also spoke the words and pseudowords used as targets in the experiment. These pitch contours were resynthesized using a format spectrum corresponding to the resonance of the human vocal tract to give them a voice-like quality while eliminating vowel or segment content. Another set of primes consisted of the musical representations of the four tones as they are played in *suab raj nplaim* (Poss 2005). Words with high-level tones and mid-rising tones are both performed on the highest pitch level in *suab raj nplaim*, so the same recording was used for both. The fundamental frequencies of the instrumental primes fell in a range between 274 to 450Hz while the vocal primes fell in a range between 100 and 150Hz. A control prime was produced using the amplitude envelope of a spoken word resynthesized with white noise so that it would not have any defined pitch. The primes were digitally edited to be of similar length (402 ms) and intensity (ca. -22 dBFS).
Figure 4.1. Instrumental (left) and vocal (right) primes with melodic contours corresponding to the four Hmong tones used in the experiment (left-to-right: high level, mid-rising, low level, and high falling. Top: waveforms; bottom: spectrograms with pitch (blue) and intensity (yellow) curves.

The target sounds consisted of 80 monosyllabic common nouns, verbs, and adjectives as well as 80 pseudowords recorded by a native male Hmong speaker. There is no research available on word frequency in any Hmong dialect, but the Southeast Asia Community Resource Center published a White Hmong primer that was based on what they considered to be high-frequency words (Lewis & Vang 1999). This list, as well as my own experience learning White Hmong and communicating with speakers of Hmong, were used to select the final word list. All words and pseudowords began with a consonant or a consonant cluster and ended with a vowel, although some nasalized vowels can be analyzed as ending with the consonant [ŋ]. The average file length for pseudowords was 587 ms whereas the average file length for words was 554 ms (33 ms difference). This difference may be due to the fact that more of the pseudowords began with consonant clusters than
the words. It is also possible that the speaker also spoke more slowly for pseudowords since he was unfamiliar with them.

4.3: Subjects

The 26 Hmong Americans (10 women and 16 men) who participated in the study ranged in age from 19 to 60 with a mean age of 38. 19 of the 26 subjects were born and raised in Southeast Asia (Laos or Thailand) while the rest were either born in the United States or had spent the majority of their childhood here. None of the subjects who grew up in the United States reported that they played traditional Hmong musical instruments. This is in line with what I observed in my previous research. Of the remaining 19, 10 reported that they played at least one of the instruments traditional to Hmong culture. The youngest of these was 23, but overall they were older than those who did not report playing a Hmong instrument (mean age: 48). Nine of the musicians were from Laos and one was born and raised in the Ban Vinai refugee camp in Thailand.

In this instance, consonant clusters were those that contained more than one phoneme. For example, syllables that began with f, m, or r were considered simple while syllables that began with hm, ml, or ntch were clusters. An attempt was made to balance the onset types between words and pseudowords, but the results were not perfect. Pseudowords were created by altering the tone or vowel of common words or by producing syllables with combinations of onsets and vowels that do not occur in the language. 35 of the 80 words selected for the experiment (44%) had consonant cluster onsets. Finding pseudowords with simple onsets proved challenging and in the end 56 of the 80 pseudowords had consonant cluster onsets (70%).
4.4: Procedure

Each target was presented three times with a different prime type: matching, non-matching, and control. Half of the targets were presented using matching and non-matching primes based on instrumental sounds (i.e. instrumental primes) and the other half were presented using primes based on spoken lexical tones (i.e. vocal primes) (Figure 4.2).

![Diagram](image)

**Figure 4.2.** The organization of the experimental materials. Note that the diagram only provides an example of the subdivisions at each level.

Stimulus pairs were created by combining primes and targets with an SOA of 250 ms. Stimuli were presented in three runs separated by short breaks. The stimulus
pairs were arranged so that target syllables only occurred once in each run and that the prime types were balanced (control, match, non-match, instrumental, and vocal). Presentation of the stimuli was controlled using DMDX (Forster & Forster 2003). Subjects listened to primes and targets on headphones and were instructed to speak the targets out loud as quickly and accurately as possible. A microphone attached to the computer was used to record their responses and DMDX measured the response latency (response time). If no response was made within 2000ms, the program went on to the next item. The responses were checked for errors in consonant, tone, vowel, or problems that would have affected the measurement of the response latency, e.g. extraneous oral sounds, coughing, etc. If one of the three presentations of a target had a problematic response, all three responses to the targets were excluded from analysis.

After the explanation of the experiment, each subject did a practice run of 10 items that did not occur in the remainder of the test. This was done to make sure the computer and microphone were in working order and that the subject understood the experimental task.

4.5: Results

Analysis of the response times using repeated-measure ANOVA shows a significant effect of the prime type. By-subject analysis with prime condition as main factor reached significance: $F(2,18) = 31.15; p < 0.0001$. The mean response times were 713.77 ms for the control condition, 695.75 ms for the matching condition, and
698.35 ms for the non-matching condition (Figure 4.3). Post-hoc analysis showed that the difference in response times (RT) between the control condition and the matching condition (19 ms) and the difference between the control condition and non-matching condition (16 ms) were highly significant (p < .0001). The difference between the matching and non-matching condition (3 ms) was not significant (Figure 4.3).

Figure 4.3. Response times for the control, match, and non-match conditions for pseudowords and words.
The by-subject analysis with lexical status (word or pseudoword) did not reach significance: as F(2,18) = 2.73; p = 0.11. There was also no significant interaction between the prime condition and lexical status (p = 0.17). While these results did not reach significance, there was a large difference between mean response times to words and pseudowords (68 ms). The source of this difference may be the difficulty in speaking novel pseudowords. Even the native speaker, a teacher of White Hmong, who recorded the stimuli produced recordings that were, on average, 33 ms longer for pseudowords. This difference may be compounded by the higher proportion of pseudowords that began with consonant clusters (70%) compared to the proportion of words that began with consonant clusters (44%). For both words and pseudowords, responses to targets beginning with consonant clusters took longer than responses that had a single consonant onset. There were 909 responses to pseudowords beginning with a consonant clusters for an average response time of 749 ms while the 408 responses to pseudowords beginning with a single consonant averaged 685 ms, a difference of 64 ms. For words, the 635 responses to consonant cluster onsets averaged 692 ms while responses to single consonants averaged 635 ms, a difference of 57 ms. While the difference in response time to syllables with consonant clusters is similar between words and pseudowords, there were many more responses to pseudowords with consonant cluster onsets than words with consonant cluster onsets (909 versus 635). This difference also likely contributes to the overall difference between response times to words and pseudowords.
Since the targets primed by instrumental and vocal primes were different, direct comparison of the reaction times for the two conditions is not possible. To investigate the facilitation effects of these primes, the mean differences between the control and match (cl – ma) or non-match (cl – nm) were used, resulting in variables that have negative values if the prime condition produces facilitation. This difference between the reaction times for the two conditions is comparable. Using these measures, the effect of instrumental or vocal prime type is also significant. The match condition of the instrumental primes produces a facilitation of 11.37 ms and the match condition of the vocal primes produces a facilitation of 24.51 ms (Figure 4.4). The difference between the two is significant (F(1,18) = 15.89; p = 0.003). The difference in the facilitation effect for responses to words and pseudowords is 8.38 ms and does not reach significance (F(1,18) = 3.02; p = 0.091) (Figure 4.4).
Figure 4.4. Facilitation in the match condition (RTco – RTma) for both instrumental and vocal primes split by pseudowords and words.

The difference between response times to the non-match condition versus the control condition is also significant. Instrumental primes produce a facilitation of 11.37 ms, and vocal primes produce a facilitation of 24.51 ms (Figure 4.5). The difference of 13.26 ms is significant ($F(1,18) = 13.28; p = 0.0008$). The difference between the facilitation effects for responses to words and pseudo words is 7.89 ms and does not reach significance ($F(1,18) = 2.13; p = 0.15$) (Figure 4.5).
4.6: Discussion

The intent of this experiment was to see if a pitch from a musical instrument could activate words with a corresponding lexical tone in the mental lexicon, or facilitate their activation. It was thought that a pitch contour mapped to a lexical tone might speed a response to that word. A facilitatory effect was found for matching primes, but it was not limited to words. Also, there was facilitation for primes that did not match the tonal content of the targets. This does not necessarily invalidate the original hypothesis, but it indicates that the pitch primes could affect speech processing through a non-lexical pathway. Such a pathway, which works in parallel...
with a lexical pathway, has been described in previous research (Coltheart et al. 1993; Hanley et al. 2004). Poeppel & Hickok (2000, 2004) have proposed a speech-processing framework including a processing stream that maps sound on to meaning and a stream that maps sound onto articulatory-based representations. They point to evidence from studies of aphasics who maintain the ability to comprehend words even when they perform poorly on sub-lexical speech perception tasks (e.g. discriminating between different phonemes or syllables), as well as neuro-imaging studies that show activation in particular dorsal regions of the brain when subjects are asked to perform sub-lexical tasks on auditorily presented speech. Since lexical processing is not necessary for repeating syllables in a shadowing task, the pitch primes might be acting upon the pre-motor areas of the brain involved in speech processing, producing an advantage for spoken responses. Since this would bypass the mental lexicon, primes could affect both words and pseudowords. This has been found in an fMRI study by Peschke et al. (2009), which demonstrated that the areas of the dorsal processing stream described by Poeppel & Hickok are activated when German speakers shadow pseudowords. This result does not rule out the possibility that there is interaction between pitch contours and lexical selection, but there is no evidence of it here. It is somewhat surprising that even instrumental sounds should produce facilitation, however. The vocal primes had the spectral characteristics of the human voice, which would have been processed as speech and activated the motor areas of the brain associated with speaking. The instrumental primes, in contrast, were unaltered from how they are
performed on the *raj nplaim*, and subjects knew that they were sounds from this musical instrument rather than a human voice.

Other studies have found that musical sounds activate areas of the brain related to speaking. Hickok et al. (2003) performed an fMRI study in which subjects listened to nonsense words or sequences of pitches (what they call “melodic tonal sequences”) and then mentally rehearsed them. Compared to simply listening to the stimuli, mentally humming or repeating the nonsense words caused increased activation in dorsal areas of the brain connected to speaking. Of particular interest is a left-lateralized region at the parietal-temporal boundary in the posterior Sylvian fissure called area Spt, which has been associated with integrating auditory representations of speech in the superior temporal lobe with motor representations of speech in the frontal lobe. The activation of this area by both spoken and instrumental stimuli suggests that it is not specific to speech, although certain parts of area Spt were more strongly activated by either speech or music. These differences hold out the possibility that while Spt is an area that supports auditory-motor integration in general, subfields within it might be weighted towards different stimuli. Pa and Hickok (2008) have since done more research on area Spt. In an fMRI study of skilled pianists, they found that it was activated more strongly when the subjects mentally hummed a melody than when they imagined playing it on a piano. This indicates that area Spt is specific to the vocal tract motor system rather than other systems.
These studies indicate that it is possible for both spoken and instrumental stimuli to active areas of the brain associated with auditory-motor integration for the vocal tract. The vocal and instrumental primes elicited different effect sizes indicating that even if they both result in rehearsal, there are likely differences in how they are processed. This could be related to the finding from Hickok et al. (2003) that the spoken and instrumental stimuli caused activations in slightly different places in area Spt. In their experiment, however, the stimuli were quite different—nonsense words versus melodies. The primary difference between the primes the in the Hmong language experiment is their contrasting timbre. This suggests that processing difference is based, at least in part, on the spectral characteristics of the sound. For example, it could be that the vocal primes caused a greater facilitation because their similarity to the sound of the human vocal tract produces a stronger activation.

Indeed, several studies have found activation in the posterior temporal lobe for both instrumental and vocal stimuli. Belin et al. (2000) found that vocal and non-vocal stimuli (including the sounds of instruments) activated similar parts of the superior temporal sulcus, although vocal sounds produced a stronger pattern of activation. Schön et al. (2005) performed several different types of experiments (fMRI, EEG, and behavioral) and found consistent overlaps in the processing of speech and singing in the temporal lobe. The fMRI data demonstrated that the stimuli activated similar parts of the brain and caused ERPs with similar timecourses. It also showed that singing and speech caused different levels of activation. This evidence supports
the view that speech and music share neural networks, although certain areas may be more strongly activated by speech or vocal sounds.

A facilitatory effect at the premotor does not help to explain how Hmong people understand messages played on the raj. The effect observed here obtained for both matching and non-matching primes, in contrast first experiment of the present study in which subjects consistently identified words that matched the melodic contours of phrases. The pitch primes could have a lexical effect that is not revealed by the experimental paradigm, including a difference in processing for the match and non-match condition. Subsequent experiments are discussed in the next chapter, which suggest this is indeed the case.

The question of whether or not the practice of speech surrogacy influences the facilitatory effect of the pitch-contour primes cannot be answered definitively in this experiment, either. It was not possible to directly compare the response times of Hmong people who did or did not practice speech surrogacy since the two groups were correlated with age: younger people were less likely to play raj and older people were more likely. Both groups did show a similar pattern of facilitation, in which words primed by the control sound were responded to significantly slower than those with pitch primes. This supports the idea that the effect observed is not related to the practice of speech surrogacy, but this cannot be known for certain for the time being.
Chapter 5: New directions in research

5.1: The cross-cultural study of pitch priming

While the impetus for this project was the practice of Hmong speech surrogacy, it became clear that it must be situated cross-culturally. The results of the priming study had to be compared against the results obtained from contrasting populations. The Hmong-speaking subjects showed facilitated responses to targets following pitch primes. Would the same be true for speakers of other tone languages? As discussed in Chapter 4, since there was a confound between the age the Hmong subjects and their ability to understand message played on the raj, it was not possible to attribute differences between the groups to the practice of speech surrogacy. Both groups responded in a similar pattern, although younger subjects responded more quickly, which suggests that the observed effect of the pitch primes is not dependent on speech surrogate experience. Testing subjects from other language backgrounds without a speech surrogate would strengthen this position or could indicate that it needs to be reconsidered. Even if speech surrogacy practice is not relevant for the effect, what about the effect of speaking a tonal language? Previous studies demonstrate that language experience affects processing in the auditory cortex (Salmelin et al. 1999; Vihla 2002), so it is possible that speakers of
other languages would respond differently. The advisor to this project, Udo Will, and Tsun-Hui Hung, a graduate student in ethnomusicology, worked with me to create subsequent priming experiments to study the influence of language experience on pitch perception and its effect on speaking. We selected two contrasting language groups that do not practice a widespread speech surrogate: Mandarin Chinese, a tonal language, and English, a non-tonal language.

The priming experiment discussed in Chapter 4 investigated if pitches could activate words in the mental lexicon that had a related lexical tone. While this effect was not found, the results indicated that pitch primes facilitate spoken responses to words and pseudowords. The fact that both words and pseudowords were affected suggests that this facilitation is not related to lexical processing, but works via another pathway. It was suggested that the pitch primes could be activating areas of the pre-motor cortex associated with planning the gestures of speech, resulting in speeded responses. The research group created a similar experiment to the Hmong priming study. The results are currently being prepared for publication, but are summarized here.

First, speakers of Mandarin Chinese were tested. Words and pseudowords were primed using instrumental and vocal sounds. Mandarin has four lexical tones (Chao 1968): high-level, rising, low-dipping, and high-falling. Vocal primes were made by extracting pitch contours from the four Mandarin tones spoken by a native female speaker who also recorded the targets. These were resynthesized with a formant spectrum that corresponds to the human vocal tract so that the primes would have
the timbre of a human voice, but not a distinct vowel. Since there is no widespread speech surrogate in Chinese culture, the instrumental primes were created by performing the pitch contours of the four lexical tones on the Chinese vertical flute, xiao. The fundamental frequencies of the instrumental primes were selected that were near the second harmonic of the vocal primes, which made them perceptually similar. A control prime was generated by convolving the amplitude envelope of a spoken syllable with white noise. The experimental materials were arranged in a slightly different way than in the Hmong experiment. Targets were presented with control, matching, and non-matching primes, but to limit the number of times that the targets were repeated, a different set of targets were used for the non-matching condition. 80 targets (40 words and 40 pseudowords) that were presented with the control and matching primes (40 vocal primes and 40 instrumental primes) for a total of 160 presentations. An additional 80 targets (40 words and 40 pseudowords were presented with non-matching primes only (40 vocal primes and 40 instrumental primes). These materials were presented in two lists in which the targets only occurred once. The SOA was 250ms as in the original experiment. 16 native speakers of Mandarin were tested, none of whom had any formal music training.

The results showed a similar response pattern to the Hmong-language experiment in the control and match conditions. Subjects responded significantly faster to targets, both words and pseudowords, following match primes (mean RT 916 ms) compared to those following the control primes (951 ms), a difference of 35 ms (See
Table 5.1). Pseudowords were responded to more slowly than words, as in the Hmong language experiment, but the difference was a smaller difference, 36 ms vs. 68 ms. Vocal and instrumental primes were paired with different sets of targets, so the response times are not directly comparable, but both prime types produced significant facilitation when comparing the match and control conditions (51 ms and 23 ms, respectively). This follows the pattern found in the Hmong language experiment in which the vocal primes had a stronger priming effect than the instrumental primes. An important difference between the results of the Hmong and Mandarin language experiments is that responses to the non-matching primes did not produce facilitation in the Mandarin language experiment. The mean RTs were 951 ms for the control condition and 950 ms for the non-match condition. In the Hmong language experiment there was no significant difference in the facilitation produced by the match and non-match primes.

![Figure 5.1. Interaction plot for lexical status and prime condition. p: pseudowords, w: words, c: control, m: matching, n: nonmatching, rt: reaction time.](image-url)
The similarities in the two experiments demonstrate that the effect of the pitch primes is not limited to Hmong culture. The differences found in responses in the non-match condition are difficult to explain. The difference could be due to the fact that different targets were used for the match and non-match conditions in the Mandarin experiment, so they are not directly comparable. A more likely explanation would be differences between Hmong and Mandarin language, the most obvious being the number of tones (eight versus four). Further study would be needed to examine the source of this variation, but recent EEG experiments performed by Udo Will and Tsun-Hui Hung using the materials from the Mandarin-language experiment indicate that the match and non-match primes result in different event-related potentials for the targets. Although the results are preliminary, they indicate that relationship of the prime to the target might affect speech processing, at least for speakers of Mandarin. It is possible that the relationship of prime and target also affect speech processing for speakers of Hmong, but that these effects are not apparent in the behavioral (i.e. reaction time) results because the different processes might operate over a similar time course.

While the new experiment shed some light on the original effect observed in the Hmong language study, other experiments were pursued to answer additional questions. In the next experiment, we examined the effect of a shorter SOA. As discussed in Chapter 4, Lee (2007) found that minimal tone pairs in a mediated priming paradigm with a shadowing task produced facilitated responses at an SOA of 50 ms. Facilitation was not found at an SOA of 250 ms, which led Lee to conclude
that tone is processed more slowly than segmental information. This allows for early form priming of words that match segmentally but not in tone. We wanted to investigate the time course of the priming effect found in the Mandarin language experiment and to find out if words and pseudowords would be affected in different ways.

The second experiment used the same experimental materials as the original Mandarin language experiment, but the SOA was shortened to 60 ms. The results only showed a significant priming effect for pseudowords (mean RTs for the control and match conditions 929 ms and 901 ms, respectively) (see Figure 5.2). Responses to words showed almost no difference between the control and match conditions (881 ms and 879 ms).

---

19 Fifteen subjects were tested, all of whom were native speakers of Mandarin and had not participated in the first experiment.
This result demonstrates that responses to words and pseudowords are affected by pitch contour primes in very different ways—a conclusion that was not apparent in the first two experiments. In the Hmong and Mandarin language experiments with an SOA of 250 ms, responses to words and pseudowords showed similar patterns of facilitation. In this experiment with an SOA of 60 ms, the facilitation effect was found for pseudowords, but not words. As noted above, Lee (2007) found that this ability of tone to constrain lexical activation was lost when the SOA was shortened from 250 ms to 50 ms. This suggests that while tone affects lexical processing early on, it operates on a different time scale than segmental information. The current experiment demonstrates a similar response pattern. Since the change is limited to words, it must be the case that pitch primes affect the processing of pseudowords at
a later stage. It could still be the case that the primes facilitate spoken responses and that this effect obscures the lexical effect found at a shorter SOA.

Subsequent experiments were performed with same materials using English-speaking subjects. This was done to examine the role of language experience on the facilitation effects found in the previous experiments. Was the facilitatory effect found for speakers of Hmong and Mandarin related to the practice of a tonal language? It may seem counterintuitive to use the Mandarin-language primes and targets with subjects who cannot understand them, but since the facilitation is not necessarily related to lexical status it seemed likely that the effect would remain. Using the same materials allowed for direct comparison between the two groups: speakers of a tonal language and speakers of a non-tonal language.

The results and analysis are currently being prepared for publication, but the early indication is that English speakers show a similar pattern of responses to those of Mandarin speakers at an SOA of 250 ms, indicating that pitch primes are able to facilitate spoken responses regardless of language experience. There are differences in the timing of responses. The responses of the English speakers were slower and showed a smaller degree of facilitation, which suggests that language experience influences the processing of the primes to some extent. With an SOA of 60 ms, there was no evidence of facilitation. This contrasts with the Chinese subjects who showed facilitation for pseudowords even at this short SOA, indicating that there is a difference in the speed with which tone is processed and available for processing pseudowords. Previous studies demonstrate that language experience affects
auditory processing in the auditory cortex (Salmelin et al. 1999; Vihla 2002) and the pitch processing of lexical tones (Gandour et al. 2000; Hsieh 2001; Wang et al. 2004). The results from this experiment could be related to these differences between speakers of tonal and non-tonal languages.

These experiments might seem only tangentially related to the research presented in the previous chapters, but they are important for a number of reasons. As stated above, these experiments come directly out of the original line of inquiry: how do Hmong musicians understand words played on the *raj*? The processing of pitch and speech are implicated in this system of communication, requiring a better understanding of how pitch, lexical tone, and the mental lexicon interact. These experiments support and expand upon the results of the original priming experiment: word repetition is facilitated by pitch contour primes for speakers of both tonal and non-tonal languages. Although the effect is not directly linked to the practice of speech surrogacy, speakers of tone languages show greater facilitation. The results of at least one of the experiments (50 ms SOA for speakers of Mandarin) indicate that the pitch primes do have a lexical effect. The implication of this for our understanding of Hmong speech surrogacy will have to be developed in future studies.

5.2: Formulaic language on the Hmong *raj*

The study of phrases played on the *raj* can also be placed into a larger context.
While individual words were the primary focus of analysis in the first experiment, the results suggest that some listeners responded in terms of complete phrases. Subjects sometimes reported phrases that did not match the melodic outline of the item or the number of pitches presented. It is possible that phrases played on the *raj* constitute formulaic sequences, which Alison Wray has defined as “a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar” (2002:9). Elsewhere, Wray has argued that the mental lexicon contains “bundles of different sizes” (2008:12) like morphemes and words, but also “multiword strings, including some that are partly lexicalized frames with slots for variable material ...” Some phrases played on the *raj*, like “*Nkauj Hmoob, Nkauj Hmoob puas kho koj siab tsis kho*” (“Hmong girl, are you lonely or not?”), are played with little variation by many musicians. This particular phrase is known by almost everyone who knows how to play the *raj* or who can understand even a few of the words that played and was used as the first item in the first experiment. The variations I have found among players could be described in the following way: \([Nkauj Hmoob/([Girl from a specific clan]) [puas/yuav] kho [kho] [koj/nef] [lub] siab tsis kho. The frame of the phrase is *kho siab tsis kho*, “to be lonely or not lonely.” Around this frame, the performer has a limited choice of variations to address a specific audience (*e.g.* Hmong girl or a girl from a specific clan, singular or plural) or elaborating the expression with words that do not change the meaning (*e.g.* intensifying the verb by doubling it or adding
the classifier “lub”). Another example from the experiment is “Kuv tsis muaj kwv tsis muaj tig” (“I don’t have any brother/male relatives,” i.e. “I am an orphan/all alone in the world”). Different two-word expressions for relatives can be inserted into this frame: niam txiv (parents), kwv tig (brothers/male relatives), or viv ncaus (sisters/female relatives). The frame would be: Kuv tsis muaj [relative 1] tsis muaj [relative 2].

These two examples illustrate characteristics of formulaic sequences described by Norbert Schmitt and Ronald Carter (2004). Both have slots that allow them to be applied in different settings, and the slots have semantic constraints. These examples also demonstrate how formulaic sequences are tied to conditions of use, in this case a performance of raj on the topic of love and loneliness. Koenraad Kuiper (2004) relates such formulaic performances back to the work of Albert Lord who found parallels between the texts of Homer and contemporary singers of epic poetry (Lord 1960). Lord argued that the bards compose the poetry in real time using formulaic phrases. These phrases aid in maintaining fluency despite the lengthy performances that must adhere to a metrical framework. Kuiper goes on to describe the characteristics of oral traditions that rely on formulae, including specialized prosodic modes (e.g. chanting) and restrictions on the context of performance (2004:39), which could also describe phrases played on the raj. Wray also provides a detailed set of criteria to help researchers explain why they intuitively identify some phrases as formulaic. These criteria are not meant to prove conclusively that some phrases are formulaic, but in combination with other
methods of identification (e.g. corpus analysis) they can establish "reliable justifications" for inclusion (2008:114). Raj phrases meet many of the criteria: they are associated with specific situations, they perform a function in communication in addition to conveying the meaning of the words, they are performed as units, they are marked lexically, and they are formulated in the same way by many people.

If phrases played on the raj are formulaic and are stored and retrieved as units, it poses new possibilities for future research. For example, might predict entire phrases based on the initial melodic outline. Listeners could be primed with initial phrase fragments of varying lengths and asked to respond with what they believe the phrase is. This would help to establish how conventionalized phrases on the raj really are, depending on how the responses coincided. It would also provide information on how these formulaic units are stored in the mental lexicon in terms of length, structure, and vocabulary across the community.

5.3: Problems and possibilities in interdisciplinary research

Interdisciplinary research has become commonplace, but researchers still face a variety of challenges. Judith Becker, a pioneering ethnomusicologist who has drawn on methods from psychology to study the relationship between music and trance, has written on the barriers encountered in her own interdisciplinary work. She describes an emphasis in ethnomusicology on “intensive fieldwork" and “the integration of musical and cultural analysis” that goes along with a “distrust of either broad comparative studies or overarching theories of musical experience”
(2009:45). She notes particular points of difference between empirical methodology and the ethnomusicological mindset. “Psychological experiments with music are often concerned with very small matters. A scientist tries to limit or control multiple variables in order to be able to attain verifiable results. The ethnomusicologist is more interested in musical meaning writ large. “Meaning” always refers to that which matters to the musicians and listeners of a particular genre of music, not that which matters to the analyst.” While this is an exaggerated characterization, there is a basic truth in her statement.

I encountered this dilemma in developing the experiment testing the ability of listeners to understand raj phrases of different lengths. There is a multitude of ways to “understand” such performances: word-for-word, picking out enough words to catch the meaning, or an overall sense of the topic. It is possible to also “understand” a performance as a beautiful expression of skill or a demonstration of cultural know-how. Stripping away the context is necessary to find out how listeners use cues like pitch and articulation to pick out individual words and phrases, but it obliterates the “meanings” that might be more important for the musicians and listeners. For example, I carefully avoided saying that subjects who reported words that matched those provided by the performer got it “right.” There is no “right” or “correct” answer in understanding these performances. Even if a listener fails to grasp the topic intended by the performer, their interpretation might still be satisfying. My emphasis on the context of performance was partially responsible for
my decision to present the items in the same order to all subjects, from long to short. In this way I hoped to ease listeners into the unfamiliar experimental setting.

As I considered the types of response strategies subjects were likely to use, I became skeptical of how useful statistical analysis would be in understanding the results. Becker notes that "ethnomusicologists distrust statistics when it comes to matters so intimate and so analogical as musical expression and reception. We almost invariably follow qualitative methods and resist the lumping of humans into abstract groupings that distorts or erases their individuality." I did not necessarily distrust statistical analysis, but I was deeply concerned that I could miss important details found in the individual responses. For example, the "expression and reception" might be construed as the words intended by the performer and the words understood by the listener. It is simple to measure if the response of a subject matches that of the performer or whether the response matches in terms of pitch or onset. But what about non-matching responses that appear to be related to what the performer originally intended? For example, one subject reported "leej vi leej ncaus" ("sisters") for "muaj vi muaj ncaus" ("to have sisters"). The meaning is similar to what the performer intended and nothing about the pitch outline or onset types would favor one interpretation over the other, yet it would result in a "non-match" in the results. As was discussed in Chapter 3, people often leave out unnecessary words when reporting what was played on the raj or what they heard in a performance. This and other issues in analyzing the data led me to choose a different path—one that blended a mathematical description of the results with a
type of detailed description of individual responses more strongly associated with ethnography.

To many ethnomusicologists, the priming study that explored the influence of pitch primes on lexical processing would not be considered music research; the pitches used are too distantly removed from the context of music making and the experimental setting is too limiting. As Becker says, it has to do with “very small matters” rather than “meaning writ large.” From my perspective, the priming study cannot be divorced from the musical context since the Hmong subjects hear the instrumental pitch primes as relating back to a musical performance. At the same time, for many listeners the experience of listening to music is inseparable from interpreting a verbal message. Performances on the Hmong raj are neither totally music nor totally speech, which is why they are interesting from the standpoint of music and language research. While the results of the priming study provide some insight into Hmong culture, they also have the potential to inform our understanding of pitch and lexical processing general.

5.4: Conclusion

This entire line of inquiry would not have been possible without the original question: how do Hmong people understand words played on musical instruments? Questions like this arise out of intersections between disciplines like ethnomusicology and psycholinguistics. Most ethnomusicologists would not think to investigate speech surrogacy using methodology from cognitive science.
Psycholinguistics has yet to embrace this field—perhaps because the subject is too far removed from language as it is usually understood. Yet at a time when the study of music is proving a rich ground for the study of cognition, it makes sense to explore these areas on the boundaries. Phenomena like speech surrogates force us to rethink how we categorize our experience: music versus language, culture versus cognition. There is the potential to expose previously ignored points of overlap between these domains and create new areas of research that had not been imagined. As demonstrated in this project, and the work of other interdisciplinary scholars like Judith Becker, there are many obstacles to progress: miscommunication between researchers from different philosophical backgrounds, the need for researchers to develop skills in a variety of methodologies, and a limited audience for results that do not fit neatly into one field or another. But this project also indicates that there is hope for new approaches to the study of music and language and that what seems risky at the moment can grow into exciting new areas of research.
References


