NORTHWESTERN UNIVERSITY

The Theory of Adaptive Dispersion and Acoustic-phonetic Properties of Cross-language Lexical-tone Systems

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of linguistics

By

Jennifer Alexandra Alexander

EVANSTON, ILLINOIS

December 2010

ABSTRACT

The Theory of Adaptive Dispersion and Acoustic-phonetic Properties of Cross-language Lexical-tone Systems

Jennifer Alexandra Alexander

Lexical-tone languages use fundamental frequency (F0/pitch) to convey word meaning. About 41.8% of the world's languages use lexical tone (Maddieson, 2008), yet those systems are under-studied. I aim to increase our understanding of speech-sound inventory organization by extending to tone-systems a model of vowel-system organization, the Theory of Adaptive Dispersion (TAD) (Liljencrants and Lindblom, 1972). This is a cross-language investigation of whether and how the size of a tonal inventory affects (A) acoustic tone-space size and (B) dispersion of tone categories within the tone-space.

I compared five languages with very different tone inventories: Cantonese (3 contour, 3 level tones); Mandarin (3 contour, 1 level tone); Thai (2 contour, 3 level tones); Yoruba (3 level tones only); and Igbo (2 level tones only). Six native speakers (3 female) of each language produced 18 CV syllables in isolation, with each of his/her language's tones, six times. I measured tonal F0 across the vowel at onset, midpoint, and offglide. Tone-space size was the F0 difference in semitones (ST) between each language's highest and lowest tones. Tone dispersion was the F0 distance (ST) between two tones shared by multiple languages.

Following the TAD, I predicted that languages with larger tone inventories would have larger tone-spaces. Against expectations, tone-space size was fixed across level-tone languages at midpoint and offglide, and across contour-tone languages (except Thai) at offglide. However, within each language type (level-tone vs. contour-tone), languages with smaller tone inventories

had larger tone spaces at onset. Tone-dispersion results were also unexpected. The Cantonese mid-level tone was further dispersed from a tonal baseline than the Yoruba mid-level tone; Cantonese mid-level tone dispersion was therefore greater than theoretically necessary. The Cantonese high-level tone was also further dispersed from baseline than the Mandarin high-level tone – at midpoint and offglide only.

The TAD cannot account for these results. A follow-up analysis indicates that tone-space size differs as a function of tone-language type: level-tone and contour-tone systems may not be comparable. Another analysis plots tones in an onset F0 x offglide F0 space (following Barry and Blamey, 2004). Preliminary results indicate that the languages' tones are well-separated in this space.

ACKNOWLEDGMENTS

First and foremost, I thank my committee members, Ann R. Bradlow, Patrick C.M. Wong, and Matthew Goldrick, for their encouragement, advice, and support. I owe a special debt of gratitude to my primary advisor, Ann Bradlow. Ann, you are exactly the kind of advisor I hope someday to be. Thank you so much for everything.

I also thank the faculty and staff of the Department of Linguistics, in particular the Director of Graduate Studies, Janet Pierrehumbert, and the Department Assistant, Irene Sakk. Thanks to the NU Speech Communication and Phonatics groups for their suggestions on this project, and to my fellow students for providing intellectual and social support. Special thanks go to classmates Ross Baker, Robert Daland, Meredith Larson, and Kristin van Engen. Many thanks go to Chun Chan and Daniel Kiefer for providing technical support and Perl and Praat scripts. Thanks to the Graduate School and Department of Linguistics for their financial support.

Others deserving of great thanks include Rachel, Patrick, Kathy, Pat, Mike, Caroline, Jeremy, Elisa, James, Sara, and Midam for local friendship and support; and Tara, Judy, and Joel for support and friendship from afar.

Thanks so much to my family, especially Mom, Dad, Andj, Chris, Jeff, Grandma, Aunt Diane, Uncle Bob, Aunt Kathy, Uncle Dennis, Keith, Jess, Anna, Sam, and Nicholas; and my in-law family, Mary Beth, Joe, Jake, Meghan, Matt, Dori, Steve, Connor, and Jackson.

Finally, I thank my fiancé, Daniel, for his emotional support, neck rubs, sustenance in the form of flowers, coffee, cheese, beer, and chocolate…and for his willingness to do dishes and cook dinner while I typed. BB, I am very lucky to have you in my life.

DEDICATION

For Irene Troyanos.

Grandma, your unwavering love and support made a bigger difference in my life than I ever could have imagined. I miss you and love you.

TABLE OF CONTENTS

LISTS OF FIGURES AND TABLES

Figures

Tables

CHAPTER ONE: INTRODUCTION AND LITERATURE REVIEW

1.1. Introduction

In principle, an overarching goal of linguists is to examine and describe all languages as accurately as possible. This serves to document the complexity of the world's languages and to facilitate understanding of the complexities and range of human psycholinguistic abilities. Complete understanding of the structure and organization of linguistic systems, how they interact, and how humans process the varied information, is only possible by the thorough investigation of *all* aspects of language. Despite the fact that lexical tones are a component of about 42% of the world's languages (Maddieson, 2008), lexical-tone systems are under-studied compared to segmental contrast systems (consonants and vowels). The overarching goal of this study is to increase our understanding of speech-sound inventory organization by extending a well-studied model of vowel system organization – the Theory of Adaptive Dispersion (TAD) (Liljencrants and Lindblom, 1972) – to lexical tone systems. In particular, this is a cross-language investigation of whether and how the type and number of tones in a language's inventory (its tone inventory composition) affects (A) its acoustic tone-space size and (B) the dispersion of its tone categories within the tone space.

A key element of a comprehensive study of lexical tone systems is the judicious inclusion of tone systems and inventories that compare and contrast critical properties of tones. To this end, I examine three East Asian languages that have both contour and level tones – Cantonese (3 contour tones, 3 level tones), Mandarin (3 contour tones, 1 level tone), and Thai (2 contour tones, 3 level tones) – and two Nigerian level-tone-only languages, Yoruba (0 contour tones, 3 level tones) and Igbo (0 contour tones, 2 level tones). Such diversity facilitates examination of general principles of tone organization, via specific research questions such as: Do languages with larger tone inventories make use of a larger acoustic space than languages with smaller tone inventories? By including a range of languages I aim to provide a generalizable view of the effect of tone-inventory composition on both acoustic tone-space size and dispersion of tones within the tone space.

The upcoming sections of this chapter are organized as follows. In section 1.2., I review the TAD and how it approaches the study of the acoustics of vowel systems. In section 1.3., I review the literature on tone systems. Finally, in section 1.4., I provide a brief overview of the current study, including a description of the structure of the dissertation document.

1.2. The Theory of Adaptive Dispersion (TAD)

The main aim of the Theory of Adaptive Dispersion (TAD) (cf. Liljencrants and Lindblom, 1972; Lindblom, 1975; Lindblom, 1986) is to predict the phonetic structure of the vowel inventories of the world's languages. Crucially, the TAD evaluates the role that perceptual contrast plays in vowel systems, positing that the vowels of a given language are positioned in phonetic space in such a way as to make them highly contrastive. Certain predictions of the theory have changed over time, including the predicted distance in acoustic space for vowels to be considered maximally (or sufficiently) contrastive; assumptions regarding language-universal vs. language-specific effects on vowel dispersion and vowel space boundaries; and quantitative characteristics of the vowel space boundaries.

Liljencrants and Lindblom (1972) utilize a principle of *maximal* contrast within a universal vowel space. This universal vowel space is modeled after a typical male speaker's acoustic output of vowels according to specifications of the position and shape of the jaw, lips, tongue, and larynx as defined by Lindblom and Sundberg's (1969, 1971) articulatory model of speech production. The articulatory constraints of the model determine the range of vowel sounds producible by the vocal tract; the vowel inventory of a given language is comprised of a subset of these producible sounds. The vowels are located in a three-dimensional acoustic space defined by the first three formant frequencies (in Hz). Liljencrants and Lindblom then transform the linear frequency scale into the quasilogarithmic (mel) scale, as this more accurately reflects the manner in which the auditory system perceives sound contrasts (Fant, 1973). In a given inventory, vowels are predicted to be maximally dispersed across the vowel space, with as many vowels as possible finding equilibrium at equidistant intervals along the boundaries of the acoustic vowel space. The perceptual distance between any two vowels is calculated as being the linear distance in mel units between the points representing those vowels. For ease of visualization, Liljencrants and Lindblom redefine the vowel space using just two dimensions: F1 and F2´. F1 conveys articulatory opening and vowel height, while F2´, which is a combination of F2 and F3, conveys frontness/backness and rounding. This approach appears to reasonably successfully predict three-, four-, five-, and six-vowel inventories attested in early cross-linguistic surveys (those of Trubetzkoy, 1929; Hockett, 1955; and Sedlak, 1969). No major discrepancies exist between Liljencrants and Lindblom's computer-generated simulations and actual attested three-vowel systems. Just as predicted, attested systems usually contain what are the three most common vowels in the world's languages: the corner (point) vowels [i, a, u]. Given the range of F1/F2 values that are producible in vowels, these vowels, which are maximally distinct, can be most often distinguished from one another. Also as predicted, most attested four-vowel systems contain [i, ɛ, a, u], which are the next most common vowels cross-linguistically (Crothers, 1978; Maddieson, 1984). Minor discrepancies exist between predicted and attested five-, and six-vowel systems. Natural seven- to twelve-vowel systems had a lower number of high-vowels than was predicted by the model. Predicted seven- and eight-vowel systems lacked the attested interior mid vowels such as [ø] and exhibited four, rather than two or three, degrees of backness in the high vowels. Predicted nine-, ten-, eleven-, and twelve-vowel systems had five degrees of backness in the high vowels rather than the attested four or fewer degrees of high-vowel backness.

In order to address these discrepancies, Lindblom (1975) revised the TAD so as to give more weight to the F1 dimension and less to the F2´ dimension. This is motivated by the observation that F1 is favored in vowel contrasts over higher formants. Lindblom (1975) posits that vowel systems, developed so as to guarantee some amount of perceptual clarity under suboptimal acoustic conditions, would be expected to exploit F1 (height or sonority) more than other formants because F1 is more intense and is therefore more salient in noise. Predictions for seven- to nine-vowel systems are improved as a result, but they remain imperfect. Specifically, for systems of seven or more vowels, it predicts more degrees of high-vowel backness than is attested.

Lindblom (1986) revises the TAD even further, questioning the adoption of the formant-based distance measure. He takes a cue from Bernstein (1976) which found that it was not possible to describe perception of steady-state synthetic vowels solely in terms of F1, F2, and F3. Lindblom notes that, while we might suppose that spectral peaks play a significant role in determining vowel quality, there is in fact little evidence to suggest that the ear literally tracks formants and discards all other information. (Lindblom, 1986:23) Therefore, Lindblom abandons the assumption that perceptual distance parameters ought to be defined on acoustic parameters, and replaces it with distance functions more relevant to the auditory perception system. He bootstraps a model by Schroeder, Atal, and Hall (1979) where an input – the harmonic power spectrum of an arbitrary vowel – is passed through an auditory filter whose parameters are defined by psychoacoustical data on pure-tone masking. The output, an auditory spectrum, represents the effect of masking on a pure tone by that vowel. This version of the model accounts for aspects of human hearing (e.g., frequency resolution). In addition to this change, Lindblom (1986) replaces the idea of *maximal* contrast with that of *sufficient* contrast. He does so because, in his words, [l]anguages offer a rich variety of phonetic realizations for a given size and shape of vowel system…This quality variation suggests that predictions should not be restricted to the criterion of *maximal* perceptual contrast which gives one unique configuration per system of size *n*. (Lindblom, 1986:32-33) To define the notion of sufficient contrast, Lindblom has the algorithm enumerate the best subset of systems (*m*) for each *n*. He assumes that sufficient contrast operates in real systems and is invariant across languages and system sizes. Following from this is the assumption that phonetic values of vowels ought to exhibit more variation in small systems than in large ones. In putting it to the test, Lindblom finds that this model generates vowel systems sharing a number of essential characteristics with natural systems. (Lindblom, 1986:34) One notable improvement is that it is less likely to over-generate high vowels in systems with six or fewer vowels. However, the model still falls short in crucial ways: it still over-generates high vowels for systems with seven or more vowels, and its predictive powers are weaker for the substitution of the notion of sufficient contrast for that of maximal contrast.

Later, Lindblom revises and renames his theory the Hyper- and Hypoarticulation (H&H) theory (Lindblom, 1990). H&H takes into account inter-speaker and intra-speaker variation in production of phonetic targets. This modification is prompted by observations suggesting that the acoustic signal alone is not sufficient for accurate lexical access. Instead, lexical access is driven by the signal *after* it has been modulated by signal-independent information. For example, the utterance less'n twenty is a felicitous response to both the questions *How many people came to the lecture?* and *What was your homework assignment?* Despite the fact that there may be no actual signal information disambiguating the two possibilities, understanding of context allows the listener to easily perceive the intended meaning (Lindblom, 1990:143). This fact is taken as further evidence that sufficient contrast, rather than maximal contrast or signal invariance, allows speech sounds like vowel categories to be differentiated. According to this version of the theory, speech production operates within a feedback loop: in short, talkers attempt to emulate hyperarticulated (clear) speech, under the presumption that sounds in hyperarticulated speech are especially contrastive.

The next section reviews literature relevant to this study: that which tests TAD predictions about (a) the effect of sound-inventory size on acoustic-space size, and (b) the dispersion of sound-categories within the acoustic space. Note that, over the years, several studies have also investigated the accuracy of the predictions of the TAD with respect to speech-sound *perception* as well, but that literature is not reviewed here, as it is outside the scope of this study.

1.2.1. Research testing predictions of the TAD

The size of the acoustic vowel-space is positively correlated with the size of the vowel inventory

Several studies have used the TAD to motivate and test hypotheses about the vowel spaces of languages with larger vowel inventories vs. those of languages with smaller vowel inventories. One key prediction of the TAD is that languages with larger vowel inventories will have larger acoustic vowel spaces, relative to languages with smaller vowel inventories. The results of some studies have supported this notion. Jongman, Fourakis, and Sereno (1989) found that English and German, with 11 and 14 monophthongs, respectively, have more crowded vowel spaces than Greek, which has five monophthongs. That is, the vowels /i/, /a/, and /u/, which are shared among the three languages, occur in similar positions in the languages' F1 x F2 and F3 vowel spaces. However, the other German and English vowels were more peripheral than the Greek vowels. Similarly, Al-Tamimi and Ferragne (2005), compared the 5 vowels of Moroccan Arabic, the 8 vowels of Jordanian Arabic, and the 11 vowels of French as produced in three conditions: in isolation, in syllables, and in words. The authors defined the vowel-space as the Euclidean distance between point vowels [i, a, u] in an F1xF2 Bark space. They found that French > Jordanian Arabic > Moroccan Arabic in vowel space size, in all three vowel-production conditions. Similarly, Bradlow (1995) found that English (11 vowels) had an expanded vowel space relative to Spanish (5 vowels), when those vowels were produced in a closed-syllable context (vowel-space was determined by intervocalic Euclidean distances in an F1 x F2 Hz space). Finally, Flege (1989) used a glossometer to compare native English speakers' vowels with native Spanish speakers' vowels. He surmised that native English speakers, who have a more crowded vowel space, would maximize the articulatory distance between point vowels by using more extreme tongue positions than native Spanish speakers. Indeed, Flege found that English vowels were produced with a greater range of vertical tongue positions. Specifically, English /i/ and /u/ had higher tongue positions than Spanish /i/ and /u/, and English /a/ was produced with a lower tongue position than Spanish /a/. Flege suggested that the reason English speakers use more extreme tongue positions to articulate vowels than Spanish speakers is because perceptual confusions are more likely to occur in English due to its larger vowel inventory.

The studies discussed above appear to support the hypothesis that larger vowel inventories lead to larger vowel spaces. However, this hypothesis seems to not be unequivocally true. As a matter of fact, this prediction of the TAD may be one of its most problematic. For instance, Gendrot and Adda-Decker (2007) compared eight languages with differently-sized vowel inventories (English, French, German, Italian, Mandarin Chinese, Portuguese, and Spanish) in order to investigate whether these languages' acoustic vowel spaces differed as a function of inventory size. The authors determined the shape and size of the languages' vowel spaces by measuring the Euclidean distance between peripheral vowels (F1-F0 x F3-F2 on a Bark scale), and found that languages with larger vowel inventories did not have respectively expanded vowel spaces. On an even larger scale, Livijn (2000) compared the differently-sized vowel inventories of twenty-eight languages that were chosen to be as genetically and typologically varied as possible. Livijn measured the sum of the Euclidean distances between F1 and F2 (in Bark) between point vowels /i/, /a/, and /u/ and plotted them as a function of inventory size. He found that the Euclidean distances between point vowels in languages with 4-8 vowels in their inventories were comparable. In other words, the distances were expanded only in languages with 11 or more vowels.

Gendrot and Adda-Decker (2007) and Livijn (2000) appear to contradict Jongman et al. (1989), Flege (1989), and Al-Tamimi and Ferragne (2005): the latter three suggest that a larger vowel inventory leads to a more expanded vowel space, while the former two do not. The reason for this discrepancy is unclear, but may possibly be due to methodological factors (see also Bradlow, 1993 for a discussion of this issue). Jongman et al. (1989) and Flege (1989) studied vowels produced in isolation; Livijn (2000) and Gendrot and Adda-Decker (2007) examined vowels as produced in words; and Al-Tamimi and Ferragne (2005) studied vowels as produced in words, syllables, and isolation. As will be discussed in detail in chapter three, the current study employs rigorously-controlled methodology for eliciting speech sounds to be analyzed, both because methodologically-varied studies impair our ability to make generalizations and well-motivated predictions, and because evaluating multiple languages under controlled conditions maximizes our ability to plausibly compare their sound systems.

Vowels will be maximally (or sufficiently) dispersed throughout the vowel space

Another assumption of the TAD is that the vowels in a language's inventory will be maximally dispersed throughout the vowel space (or sufficiently dispersed, in later versions). The literature on this topic does not consistently support this notion, however. On the one hand, Disner (1984) reported that about 96% of the 317 languages documented in UCLA Phonological Segment Inventory Database (UPSID) (Maddieson, 1984), which is based on transcribed data (as opposed to acoustic measurements of data), have vowel systems that contain vowels that approach even dispersion along the boundaries of the acoustic vowel space. On the other hand, Lindau and Wood (1977) report that while the closely-related languages Yoruba, Ghotuo, and Edo each have seven vowels, the vowels in Edo and Ghotuo are quite evenly dispersed across their respective vowel spaces, but those of Yoruba are less evenly dispersed. Likewise, Recasens and Espinosa (2006) compared the F1xF2 characteristics of the vowels of three dialects of Catalan (Valencian, Eastern Catalan, and Western Catalan) that each have seven vowels, as well as that of the vowels of a fourth system (Majorcan) that has the same seven vowels plus stressed /ə/. They found that the vowels of the three Catalan dialects were comparably dispersed across their respective vowel spaces (and that the vowel space of Majorcan was comparatively larger). However, intervocalic distances varied according to dialect and vowel pair, which is inconsistent with the TAD prediction that adjacent vowels will be evenly spaced in identical vowel systems. Additionally, Disner (1983) reported that the nine vowels of Swedish and the ten vowels of Danish are crowded into a small section of their respective vowel spaces, instead of being more thoroughly dispersed.

1.2.2. The TAD and consonant systems and click systems

Most work on the TAD is based on studies of vowel systems, but not all. At least one study has tested predictions of the TAD with respect to consonant inventories. De Jong and Obeng (2000) examined the typologically uncommon occurrence of simultaneous labial rounding and palatal constriction in Twi (labio-palatalization). Upon examination of distributional patterns, palatograms of the articulation of secondarily articulated consonants, and acoustic analyses, the authors conclude that labio-palatalization in Twi is the result of a historical and functional convergence of consonantal rounding and vocalic palatalization. Specifically, they argue that the principle of maximal dispersion explains the combinations of constriction location and rounding degree found in Twi labio-palatalization, in that both articulations contribute to a common acoustic function of altering the timbre of consonantal noise, thereby dispersing contrastive speech sounds further apart.

A small amount of work on TAD based on click inventories has been done as well. Miller-Ockhuizen and Sands (2000), in a study on the forward released dental-alveolar lateral click in Mangetti Dune !Xung (M.D. !Xung), determined that inclusion of this new click in the language’s click inventory ultimately causes the acoustic-phonetic characteristics of the entire click inventory to adjust. To accommodate the new click while maintaining maximal perceptual distinctiveness between it and other clicks in the inventory, M.D. !Xung speakers alter their production of one of its other clicks, the lateral alveolar click. As a result, the M.D. !Xung lateral alveolar click has a shorter burst duration than the same click in Jul’hoansi, which is a related language that lacks a forward released dental-alveolar lateral click. Because the larger contrastive set of M.D. !Xung clicks is less widely dispersed over the acoustic space than the smaller contrastive set of Jul’hoansi clicks, M.D. !Xung speakers manipulate the temporal cue of burst duration to ensure its clicks are distinct.

1.3. Tone systems

Millions of people across the globe speak a tone language as their native language; some of the more well-known tone languages include Mandarin Chinese, with 885 million speakers and Yoruba, with 20 million speakers (Yip, 2002:1). In some areas of the world, e.g., China, Central America, and sub-Saharan Africa, almost all the languages are tonal.

The first sub-section that follows is a brief discussion of the common defining characteristic of tone languages: their use of fundamental frequency variation to convey semantic meaning. The next sub-section builds upon this understanding of the role of pitch to describe tone inventories in general. The final sub-section discusses tone rules (e.g., rules for tone-tone interactions). This section relies heavily on the work of Maddieson (1978), Yip (2002), Hyman and Schuh (1974), and Hyman (2007).

1.3.1. Phonemic use of fundamental frequency (F0)

According to most sources (e.g., Yip, 2002; Hyman, 2001), the defining acoustic characteristic of a tone language is its phonemic use of fundamental frequency (F0) (pitch, in psychoacoustic/perceptual terms), meaning that tone languages use pitch changes to convey semantic contrasts at the lexical (word) level. Pitch variations in non-tone languages like English express pragmatic meaning; in English, pitch conveys affect (e.g., lower pitch, when the talker is unhappy), utterance type (e.g., declarative statement *You're a good student.* vs. interrogative *You're a good student?*), and emphasis (e.g., *I have a* ***cat****, not a dog*.).

The term tone language subsumes two types of languages: (1) lexical-tone languages like Mandarin Chinese, where pitch variation operates upon a language-specific segment (e.g., a syllable) and thereby systematically changes the meaning of the word; and (2) pitch-accent languages like Japanese, where pitch is also phonemic but may be restricted in distribution (e.g., on only one of the last two syllables of a word), the result of which is that not every word is a member of a pitch-contrastive minimal pair. This thesis concentrates on the former category, lexical-tone languages (as such, a discussion of lexical tone languages, but not of pitch-accent languages, follows).

In lexical-tone languages, tone is a suprasegmental feature, meaning that it operates above (independently of) the segment (cf. Goldsmith, 1990; Liang and van Heuven, 2004). The tone-bearing unit (TBU) is typically considered to be a single syllable (Yip, 2002) or the vowel of that syllable (see, e.g., Zhao and Jurafsky, 2007, 2009). For instance, in Mandarin Chinese, pitch changes across a syllable signal word meaning (e.g., the syllable /di/ with high level pitch means low, but the same syllable with falling intonation means ground).

In other tone languages, the distinctive pitch must appear somewhere in the word, but its exact location is variable depending on both the morphology of that word and the surrounding phonological context (Yip, 2002:1). For instance, in the Bantu language Chizigula (Kenstowicz and Kisseberth, 1990), some words have a low tone across all the syllables of the word, while others have one or more syllables with a high tone. Table 1.1 is reproduced from Yip (2002), who cites Kenstowicz and Kisseberth (1990); because it can be shown that the syllables with low tones are not actually phonologically specified for tone, they are called toneless. Here, the high tone is marked with the accent mark ´, as in /é/.

Toneless verbs	**English gloss**	**H-tone verbs**	**English gloss**
ku-damaɲ-a	To do	ku-lombéz-a	To request
ku-damaɲ-iz-a	To do for [someone]	ku-lombez-éz-a	To request for [someone]
ku-damaɲ-iz-an-a	To do for each other	ku-lombez-ez-án-a	To request for each other

Table 1.1. Toneless vs. H-tone verbs (Yip, 2002)

The high tones are part of the lexical entry of verb roots such as / lombéz/ (‘to request’), and occur on the penultimate syllable of the complex verb form rather than on the verb root itself each time, but regardless, it always appears so as to distinguish high tone verbs from low (toneless) ones like /damaɲ/ ‘to do’.

Because the common thread of all lexical-tone languages is their use of pitch to convey lexical meaning, linguists generally typify the tones of a language according to their fundamental frequency characteristics. From this comes a description of the inventory of a language’s tones. As Yip (2002) states, before we can describe tonal systems, we must determine how to read them, which can be difficult considering there is no consensus on how to transcribe them. Africanists (e.g., Hyman and colleagues) traditionally use a set of accent marks (´ ¯ ` ˆ ˇ) and/or Roman letters to indicate different tones; Asianists and Meso-Americanists use digits but, for the former, 5=high and 1=low while for the latter, 5=low and 1=high. For Asianists and Meso-Americanists, two digits also are used to show the pitch at the end of the syllable. Table 1.2 is adapted in part from Yip (2002:3):

			Africa		Asia	Central Am.
high	H	acute accent	´	á	55/5	1
low	L	grave accent	`	à	11/1	5
mid	M	level accent	¯ (or unmarked)	ā, a	33/3	3
fall high to low	HF	acute + grave	ˆ	â	51	15
rise low to high	LR	grave + acute	ˇ	ǎ	15	51

Table 1.2. Tone symbols (Yip, 2002)

1.3.2. Tone inventories

Level tones

Maddieson (1978) observes that while many phonetically distinguishable levels of pitch are possible in speech, no known language makes a phonological contrast of more than five tone levels. According to Maddieson, several languages have five contrastive level tones, including African languages Dan (Béarth and Zemp, 1967) and Ngamambo (Asongwed and Hyman, 1976), Asian languages Black Miao (data from F.K. Li, cited in Chang, 1953) and Tahua Yao (Chang, 1953), and American languages Ticuna (Anderson, 1959) and Usila Chinantec (Rensch, 1968). Four-level tone languages include African languages Mambila (cf. Connell, 2000) and Igede (Bergman, 1971), Asian languages Po-ai (Li, 1965) and Yay (Gedney, 1965), and American languages Chatino (Upson, 1968) and Ojitlan Chinantec (Rensch, 1968). Five- and four-level tone languages are relatively rare, however. Languages with three level tones are much more common, and those with two level tones are the most frequently encountered type of tone language (Maddieson, 1978:338). Examples of languages with three level tones include the African language Yoruba (Hombert, 1976) and the Asian language Thai (Erickson, 1974). Languages with two level tones include the African language Zulu (Cope, 1959) and the American language Navajo (Hoijer, 1945).

Table 1.3, adapted from Maddieson (1978:339), shows the F0 of each tone in an illustrative sample of two-, three- and four-level tone languages. The numerical values indicate the difference, in Hz, between the lowest tone in each system and its other tones.

	Two levels		Three levels			Four levels
	Siswati	**Kiowa**	**Yoruba**	**Thai**	**Taiwanese**	**Toura**
						+50
			+52	+28	+32	+30
	+18	+22	+27	+16	+18	+10
Lowest tone	+0	+0	+0	+0	+0	+0

Table 1.3. Pitch intervals between tones in languages with different numbers of level tones

Note first that the F0 of the highest tone of one language is not equivalent, nor even necessarily close, to that of another. Maddieson argues that, while it is *possible* for a two-level tone language to have its tones at the extremes of the pitch range, it is not *probable* for this to occur. Extra-high tones and extra-low tones do not normally occur unless there are other tones in between. For instance, systems with three level tones most frequently contain a mid-level tone along with high and low level tones (e.g., Yoruba). (For a discussion on tone-markedness constraints and their effect on tone-inventory composition, see Maddieson, 1978).

Pike (1948) suggests that, relative to a language with fewer tone levels, a language with more levels would be expected to (a) occupy a greater overall pitch range, and (b) have a smaller pitch difference between tone levels. As Maddieson points out, however, Table 1.3 shows that the tones of languages with more levels can occupy a *smaller* overall pitch range than the tones of languages with fewer levels. For instance, Toura (4 levels) occupies a 50 Hz space, while Yoruba (3 levels) occupies a 52 Hz space. Additionally, Table 1.3 indicates that the tones of languages with a greater number of tone levels are not necessarily separated by smaller pitch intervals than the tones of languages with fewer tone levels. For example, the pitch difference between the lowest and next-highest level tones in Siswati (with 2 levels), Thai (3 levels), and Taiwanese (3 levels) is *smaller* than the pitch difference between Toura's highest two tones.

Finally, note that the tones in languages with three or more levels are not equivalently separated (e.g., Toura's lowest and next-highest tones are separated by 10 Hz, while its highest and next-highest tones are separated by 20 Hz). These data may provide a first indication that tones are not made sufficiently distinctive by overall F0 differences alone. This point will become particularly important in later chapters of this dissertation, when I test hypotheses and predictions of the TAD with regard to cross-language tone systems.

The above observations notwithstanding, it is important to note that the trends shown in Table 1.3 might well be inconclusive, because the methods used to collect these data vary considerably. The Siswati data (Goldstein, 1976) reflect an average of peak F0 in the first syllable of two repetitions of eight words balanced for vowels and initial consonants but with contrastive tones, produced by a female talker. It is a reasonably rigorous and methodologically well-controlled study, but has a small *n* and is therefore limited in power. The Kiowa data (Sivertsen, 1956) reflect an average of two repetitions of a tonal minimal pair in identical (utterance-initial) environments by a male talker. The Yoruba data (Hombert, 1976) reflect the central point of tones measured from diagrams representing averages of 35 monosyllabic words produced by two male talkers. The Thai data (Erickson, 1974) reflect measurements from diagrams indicating averages of the central point of each of several tones by a male talker; number of tokens is not reported. The Taiwanese data (Zee, 1978) reflect fifteen tokens of each of the three level tones produced by each of two male talkers. Béarth and Zemp's (1967) data on Toura tones, for instance, reflect averages from several hundred utterances by one male talker, and the study omits information about the sentence frame used and specific tokens measured. The methodological variation seen in studies on lexical tone such as those reviewed above limits the extent to which results and conclusions from one study can be generalized as applicable to other languages. Once again, the current study uses controlled methodology to avoid such pitfalls.

Contour tones

Lexical-tone languages may also contain contour tones, wherein the pitch changes across the tone-bearing unit (TBU). Authorities on tone systems, including Maddieson (1978) and Yip (2002) hold that contour tones are additions to a level-tone inventory. That is, if a language has contour tones, it must also have at least one level tone. This typification of contour-tone languages is widely-accepted, and is reflected in the literature and in the current study (but see Pike, 1948; Newman, 1986; and Ray, 1967; they suggest that some contour-tone-only systems exist). A number of languages have more level tones than contour tones. For example, Yay has four level tones, one rising tone, and one falling tone (Maddieson 1978:345). Languages with two level tones and one contour tone (e.g., Zulu, with two level tones and one falling tone) are very common. Some languages have the same number of level and contour tones, e.g., Central Monpa (das Gupta, 1968), which has a high level tone and a rising tone. Other languages have fewer level tones than contour tones, e.g., Muong (Barker, 1966) which has two level tones, two rising tones, and one falling tone.

Many languages have more than one type of contour tone. Maddieson (1978:346) further stipulates that a language with complex (bidirectional, e.g., dipping/falling-rising) contour tones also has simple (e.g., rising) contours. For instance, Mandarin Chinese has, in addition to a high level tone, a rising tone, a falling tone, and a dipping tone. According to Maddieson (1978), Mandarin dialects alone contain 335 more falling tones than rising tones, which suggests that falling tones are more common than rising tones. However, this is not always the case; the Wu dialect of Mandarin has more falling tones than rising tones, and the Cantonese dialect Yüeh has the same number of falling as rising tones (Cheng, 1973).

1.4. Prior work on acoustic tone spaces and tone dispersion

To my knowledge, only three studies have attempted to quantify the acoustic tone space and the dispersion of tones therein. These are discussed in some detail here because their methods in particular inform the methods used in the current study. Zhao and Jurafsky (2007, 2009), examined Cantonese tone dispersion in plain vs. Lombard speech (speech produced in a noisy environment) and in high- vs. low-frequency words. They measured the F0 in Hz of vowels (the TBU) in CV and CVC monosyllables at ten equidistant timepoints *k* along the tonal trajectory. Timepoint *k*1 was subsequently excluded from analysis because the F0 of the initial vocalic segment can be perturbed by the preceding consonant (Hombert, Ohala, & Ewan, 1979). The authors converted the tonal F0 values to semitones (ST), because this psychoacoustic scale more accurately reflects listeners' intuitions about intonational equivalence (Nolan, 2007). For this reason, and discussed in more detail in later chapters, the ST scale is adopted in the current study as well. For each talker, the authors defined tone-space dispersion as the mean Euclidean distance of individual tones from his/her centroid; the centroid was defined at each timepoint *k* as the mean F0 at that *k*, averaged across all productions of a given tone. The distance between any two tones was defined as the summed Euclidean distance between their F0s at all points *k*. Zhao and Jurafsky found that ambient noise and lexical frequency both influence tone production. All Lombard-condition tones were produced with a comparatively higher F0. Additionally, low-frequency words with mid-range (mid-level or mid-rising) tones were produced with higher F0 than high-frequency words. The F0 trajectories of low-frequency words' F0 trajectories were also further dispersed from centroid in plain speech. These results indicate that talkers can and do alter aspects of tone production to increase tone contrastiveness. In turn, these results support the Hyperarticulation and Hypoarticulation (H&H) model (Lindblom, 1990)'s hypothesis that speakers will produce strengthened phonetic forms to counteract comprehension difficulties that can arise under certain conditions (e.g., in Lombard speech and in low-frequency words).

Barry and Blamey (2004) compared Cantonese tone productions by normally-hearing adults, normally-hearing children, and cochlear-implanted children. Citation-form tones were elicited via a picture-naming task involving 15 presentations of each of the six tone types on various (unreported) syllables, for a total of 90 items per participant. The authors plotted the tone productions in an F0 offglide x F0 onset space (in Hz), and chose this method because it captures multiple acoustic dimensions that affect listeners' perceptual judgments about tone, including average pitch, pitch direction, extreme endpoint, and slope. The tone space for each participant was calculated as ellipses surrounding the distribution of points around the mean for each of the six tones. The tonal space for children with cochlear implants was smaller than it was for normally hearing children and adults, which suggests there was little or no clear differentiation among implanted childrens' tones; it also suggests that, for normally hearing speakers, there is a direct relationship between the spread of pitch used for each tone type and the size of the tonal space. Normally hearing children also had significantly larger tonal ellipse areas than implant users, which indicates that they have a greater spread of pitch usage for each tone. The three groups of talkers in this study were clearly differentiable on observations of the locations of the F0 onset x F0 offglide points, and the degree of differentiation of the ellipses, within the tonal space. This approach to acoustic analysis of tone therefore enhances understanding of tone production based on auditory analyses.

1.5. The current study

1.5.1. Overview

My overall goal is to illuminate cross-linguistic tendencies in tone system organization. To that end, this study analyzes and compares the acoustic lexical-tone spaces, and dispersion of the tones within those spaces, of five languages with very different tone-inventory compositions.

Though tones can be defined as a combination of various acoustic correlates (e.g., mean F0, F0 turning point, duration), I follow Zhao and Jurafsky (2007, 2009) and examine one acoustic correlate across the languages: mean overall F0 across the tonal trajectory. The *tone-space size* of a language is defined as its tonal pitch range, averaged across talkers, measured at three points along the tonal trajectory (tonal onset, midpoint, and offglide). That is, the size of the acoustic tone space is measured as the F0 difference in semitones (ST) between the mean F0 of a language's highest tone and the mean F0 of its lowest tone at those three timepoints. Comparative *degree of tonal dispersion* is the cross-language difference in the Euclidean distance from the mean F0 (ST) of a given tone relative to the mean F0 (ST) of a tonal *baseline* (a basis of comparison tone chosen for being common among, and phonetically similar in, the languages).

Following the TAD, I assume that tone categories will act as repellers in a dynamical system: each will repel the others and will find equilibrium where it is maximally distant from surrounding tone categories. Closely related to this is my TAD-based assumption that tones will be dispersed only and exactly to the degree necessary to ensure sufficient tonal contrast. A third assumption naturally follows: the distance between two adjacent tone categories in a language will equal the distance between two other adjacent tone categories.

I plan to test the following two competing hypotheses:

H1. <u>Tone spaces will be equivalent in size across languages, and degree of tonal dispersion will differ as a function of tone-inventory size</u>.

a. The size of the acoustic tone space is independent of the size of the tone inventory.
b. If tone-space size is equivalent across languages, then the degree of tonal dispersion relative to a tonal baseline will be greater in a language with fewer tones than in a language with more tones.

H2. <u>Tone spaces will differ in size as a function of tone-inventory size, and degree of tonal dispersion will be equivalent across languages</u>.

a. The size of the acoustic tone space is positively correlated with tone inventory size.
b. If a language with a larger tone inventory has an expanded tone space relative to a language with fewer tones, then the degree of tonal dispersion (relative to a tonal baseline) will be equivalent across languages.

Figures 1.1 and 1.2 are idealized illustrations of five languages' tone spaces and degree of dispersion of a given tone within the tone space (here, tone 2) relative to a tonal baseline (tone

1). Languages are named A-E. The highest tone for all languages is tone 1. The lowest tone is 6 in language A, 5 in language B, 4 in language C, 3 in language D, and 2 in language E.

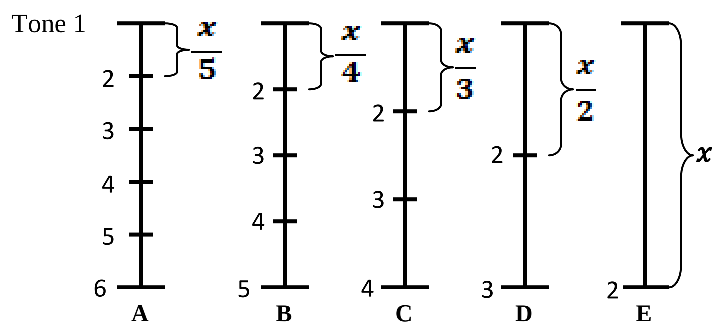


Figure 1.1. An idealized illustration of five languages' tone-space areas and degree of tonal dispersion under hypothesis H1

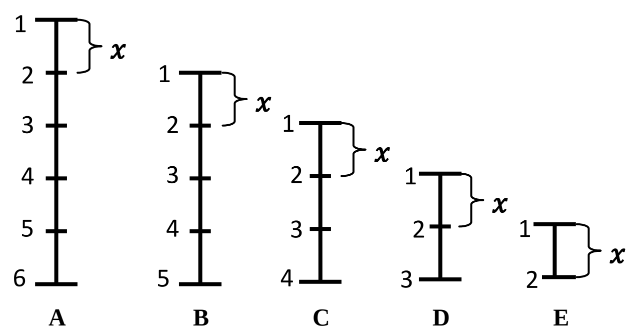


Figure 1.2. An idealized illustration of five languages' tone-space areas and degree of tonal dispersion under hypothesis H2

In chapter four, I examine cross-language tone spaces and cross-language tonal dispersion with regard to the hypotheses outlined above.

1.5.2. Choice of languages

Cantonese, Thai, Mandarin, Yoruba, and Igbo were chosen for this study for several reasons. From a theoretical standpoint, for reasons suggested earlier, it was imperative to examine languages that differed considerably with regard to the number and type of tones in their inventories. I also needed to be able to find and run participants expediently, easily, and inexpensively, and this was made possible due to the fact that a sizeable population of native speakers of each language reside in the Chicagoland area. This requirement also effectively excluded more obscure languages and dialects from consideration. Moreover, each of the languages needed to have a written transcription system that is well-known and -understood among native speakers, as participants were presented with written materials to prompt their productions of the tones. The languages also needed to have robust tone systems, that is, tone systems not in the process of major change or decay (as was apparently the case with, e.g., Burmese [Taylor, 1920]). This ensured that the speakers of each language were consistent in their understanding of their tone systems, and had sufficient metalinguistic knowledge of their languages to have the ability to produce each tone on command. Finally, I chose languages whose tone-inventory sizes were statistically common, in the hopes that the results and conclusions of this study would be reasonably generalizable to other languages.

1.5.3. Significance and Innovations

The current study is significant for multiple reasons. As noted, the vast majority of linguistic and psycholinguistic studies have concerned only segments, ignoring those 70% of languages that use tone and the more than 50% of the world's population who speak a tone language (Fromkin, 1978). This project will help to redress this balance. Also, while we clearly know a considerable amount about the individual languages' tone inventories, most prior studies on lexical tone – particularly descriptive studies conducted before the late 70s or so – exhibit considerable methodological variation, as mentioned earlier. For example, Yoruba data reported by Hombert (1976) reflect the central point of the tones as measured from diagrams representing averages of 35 monosyllabic words produced by two male talkers. Meanwhile, Thai data from Erickson (1974) reflect measurements from diagrams indicating averages of the central point of each of an unknown number of tokens of several tones by a single male talker. Such methodological non-systematicity makes it difficult to say with certainty that the results from one study can be compared with those from another. In turn, this arguably hinders our ability to make generalizations and well-motivated predictions about studies on other tone languages. One of the key aims of the current study is to evaluate the tone systems of multiple languages under conditions that are more strictly controlled, such that we may maximize our ability to plausibly compare the systems.

Furthermore, and equally importantly, no studies beyond those of Zhao and Jurafsky (2007, 2009) and Barry and Blamey (2004) have evaluated tones with respect to the predictions of the TAD, as far as I am aware. No studies have investigated cross-language tone-spaces and degrees of tonal dispersion, and none have calculated tone space and dispersion using linear mixed-effects models. The current study is therefore both innovative and serves as a contribution to the field of cross-language tone research, as its conclusions and methods can be used to motivate and inform hypotheses of future work on tone systems.

In addition, the current study is innovative with respect to its participants. It includes speakers of languages with a variety of tone systems and tone inventories. Many studies focus on the production of stimuli by one or two broadly-construed populations, e.g., tone language speakers vs. non-tone-language speakers. By more finely dividing the subject populations, the current study is expected to provide a more thorough and nuanced view of the structure and organization of cross-language tone systems.

Finally, the tone-language recordings collected at the outset of this study will be entered into a searchable database called OSCAAR (oscaar.ling.northwestern.edu) that is designed in such a way that the data contained therein may be used for – and therefore benefit – future studies. For example, a future experiment that investigates whether vowel type affects tone-space size in female Mandarin speakers might use for stimuli the female-produced Mandarin syllables produced for this study.

1.5.4. Structure of the thesis

In chapter two I present information about the languages under investigation. I first review literature on the tones of each of the five languages (Cantonese, Thai, Mandarin, Yoruba, and Igbo). I then present and describe my data on the acoustic realizations of each languages' tones, then and discuss how they compare with descriptions from the literature. In chapter three I provide information about the tone recordings that were collected and used as data for this study. In particular, I describe the methods used to recruit participants and elicit the tone-bearing syllables, and the methods used to analyze the data. In chapter four I describe the linear mixed-effects regression models used to evaluate the hypotheses described earlier. I also briefly summarize and discuss the results. Finally, in chapter five, I provide a general summary and discussion, present my conclusions and alternative analyses, and suggest future work.

CHAPTER TWO: PRESENTATION OF THE LANGUAGES UNDER INVESTIGATION

2.1. Introduction

This chapter is organized as follows. In section 2.2., I review literature on the acoustic realizations of the tones of each of the five languages under investigation in this study (namely, Cantonese, Thai, Mandarin, Yoruba, and Igbo). Then in section 2.3., I present summaries of the acoustic data I collected from speakers of each of these languages and discuss how they compare with descriptions from the literature.

2.2. Literature on the tones of the five languages

Igbo

Igbo is level-tone-only language of the Niger-Congo family (Kwa group). It is spoken by more than 15 million people in southeastern Nigeria (Liberman, Schultz, Hong, and Okeke, 1993; Hyman, 1978). It has two tones, high (H) and low (L), that occur freely; it also has a mid (M) tone that only occurs following an H (or another M). Some phonologists have treated Igbo M tones as a third, distinct, tonal category (cf. Carrell, 1970; Goldsmith, 1976). However, others claim that the restricted distribution of the M tone means that Igbo M is simply an H tone that is downstepped (a common phonological process in which high tones are lowered in a stepwise fashion after a(n overt or covert) low tone (Yip, 2002:3; Clark, 1990; Liberman et al., 1993). Monosyllables such as those in this study and thus discussed here carry only one tone: H or L. (Tones in context are not discussed here, as they are outside the scope of this project; see chapter three for the structure of the methodological design.) Table 2.1, from a native Igbo-speaking language consultant hired for this study (discussed further in chapter three), illustrates the Igbo tone contrast in monosyllables.

Syllable	**Tone**	**English gloss**
dí	H	husband
dì	L	to exist

Table 2.1. Igbo lexical tone contrasts in monosyllables

Despite the fact that Igbo tones are considered level, their phonetic values are actually determined according to their targets (the highest F0 of the H tone and lowest F0 of the L tone). These targets are found at the end of the timespan of the associated tone-bearing unit (Akinlabi and Liberman, 2000:5). For instance, in the phrase Igbo yá ('he') the monosyllabic word *yá* is considered to have an H level tone, but the pitch is not uniformly high and level. Rather, the pitch rises throughout the syllable, and the peak value is near the end (*ibid.*).

Yoruba

Yoruba, like Igbo, is a level-tone-only language of the Niger-Congo family and Kwa group. It is spoken throughout Nigeria (Hyman, 1978) and has three phonemic level tones: H, M, and L (Maddieson, 1978, 1972; Akinlabi and Liberman, 2000; Orie, 2006; and others). Generally, Yoruba tones occur freely in words, leading to three potential tone patterns for monosyllables. Table 2.2., from Akinlabi and Liberman (2000:8), exemplifies the Yoruba lexical tone contrast:

Syllable	Tone	English gloss
ra	H	to disappear
ra	M	to rub
ra	L	to buy

Table 2.2. Yoruba lexical tone contrasts in monosyllables

It is worth noting that some have suggested that the M tone in Yoruba is underlyingly toneless (see Akinlabi, 1985, Pulleyblank, 1986, and Akinlabi and Liberman, 2000). The reader is referred to those papers for a discussion.

According to La Velle (1974), a linguistic constraint in Yoruba maximizes the perceptual distinctiveness of its three (H, M, and L) tones. One specific quality of this constraint serves to lower a word-final L tone so that it may be distinguished from a word-final M tone. Hombert (1976) in particular found that the onset of the final L tone is lower in pitch, displays a falling pitch contour, is shorter in duration, and is lower in amplitude, but that F0 contour was the most salient cue to Yoruba tone identification. Indeed, neither an increase in duration nor amplitude caused shifted identification judgments, but when final L tones were manipulated to have a level (as opposed to a falling) glide, listeners misidentified L-L sequences as L-M sequences and M-L sequences as M-M sequences.

<u>Mandarin Chinese</u>

Mandarin, also known as Putonghua ('the common language') is the most widely-spoken dialect of Chinese. A Sino-Tibetan language, it is spoken throughout parts of China, including Beijing, as well as (parts of) other countries such as Singapore and Indonesia. The standard dialect is spoken in Beijing. Mandarin is typically described as having four tones in its phonological inventory, including three contour tones and one level tone. Tone 1 is a high-level tone (5-5, or H); tone 2 has a high-rising or mid-high-rising (contour) tone (3-5, or R); tone 3 has a low-dipping or low-falling-rising (contour) tone (2-1-4, or FR); and tone 4 is a high-falling (contour) tone (5-1, or F) (Chao, 1948; Howie, 1976; Blicher, Diehl, and Cohen, 1990; and many others). Mandarin also has a fifth, non-phonemic, tone in unstressed syllables which is referred to as tone 0 or neutral tone (Wong, Schwartz, & Jenkins, 2005). The F0 of the neutral tone varies depending on the tone that precedes it (Shen, 1990). Because the neutral tone does not occur in isolated monosyllables, it is not under consideration in this study. Table 2.3, from Chandrasekaran, Krishnan, & Gandour (2007) and many others, exemplifies the Mandarin phonemic lexical tone contrast.

Syllable	Tone	English gloss
ma	H	mother
ma	R	hemp
ma	FR	horse
ma	F	to scold

Table 2.3. Mandarin lexical tone contrasts in monosyllables

Mandarin tones are manifested phonetically by different overall fundamental frequency values, with F0 height and F0 contour as the primary acoustic parameters (cf. Howie, 1976; Wu, 1986; Wang, Jongman, and Sereno, 2006). As mentioned earlier, the H tone has an essentially level F0 contour (Xu, 1997); the R tone has a rising contour with a slight dip 20% of the way into the vowel (Wong et al., 2005); the pitch inflection point of the FR tone occurs approximately 50% of the way into the vowel; and the F tone rises until about 20% of the way into the vowel and then falls sharply to the end of the syllable (Xu, 1997). Other phonetic correlates of Mandarin tones include syllable amplitude (Gårding, Kratochvil, Svantesson, and Zhang, 1986); the shape of the amplitude envelope (Fu, Zeng, Shannon, and Soli, 1998); voice quality (Gårding et al., 1986); and temporal properties such as overall duration, vowel duration, and Turning Point (Lin, 1965; Chuang and Hiki, 1972; Jongman and Moore, 2000; Fu and Zheng, 2000). Regarding durational differences in particular, the H and F tones tend to be shorter than the R and FR tones for isolated monosyllables (Ho, 1976; Blicher et al., 1990). Additionally, the midpoint of the FR tone and the offglide (endpoint) of the F tone are often reported to be accompanied by a glottalized voice quality (a.k.a. vocal fry or creaky voice) (Liu and Samuel, 2004). However, it is unclear whether vocal fry functions as an acoustic-perceptual cue the same way as F0 does (Francis, Ciocca, Ma, and Fenn, 2008). All that said, tonal acoustic correlates other than F0 are outside the scope of this study, and are therefore not directly investigated herein.

There is debate in the literature as to whether Mandarin contour tones are unitary contour units or compositional sequences of multiple level tone targets. According to the former, 'unitary,' theory, for instance, tone 2 would consist of a single rise, while the latter, 'compositional,' approach would posit it as being a bi-tonal sequence of a level L plus a level H (Liang and van Heuven, 2004). Wan and Jaeger (1998), Wan (1999), and Wan (2007) argue for the former ('unitary') approach, suggesting that contrastive Mandarin tones are underlyingly linked to rimes and may therefore be unitary. Those who advocate the 'compositional' view (Yip, 1991) view Mandarin tones as consisting of a sequence of two morae (sub-syllabic timing units), each of which is a TBU. I adopt the unitary approach for this study.

Overall F0 contours provide the dominant cue for tone perception (Xu, 1997; Howie, 1976), though listeners also attend to amplitude (Whalen and Xu, 1992) and duration (Fu et al., 1998; Blicher et al., 1990; Dreher and Lee, 1966). Various acoustic cues (F0, duration, etc.) are integrated functionally when native Mandarin speakers identify the tones (Gandour, 1984; Massaro, Cohen, and Tseng, 1985; Gårding et al., 1986; Blicher, et al., 1990; Shen and Lin, 1991). But again, overall pitch contour appears to be particularly important for native listeners: native listeners attend more to pitch contour than height to make judgments of tonal dissimilarity (Gandour, 1978; Gandour and Harshman, 1978). In fact, in the presence of F0 contour, the contribution of other acoustic features is negligible for tone perception (Massaro et al., 1985).

Thai

Thai, a Tai-Kadai language with two contour and three level tones, is the official national language of Thailand. Thai has M, H, and L level tones, and F and R contour tones (Gandour, 1978). Table 2.4, from Zsiga and Nitisaroj (2007:344), displays the Thai lexical tone contrast:

Syllable	**Tone**	**English gloss**
ná:	H	aunt
na:	M	rice field
nà:	L	custard apple
nâ:	F	face
nǎ:	R	thick

Table 2.4. Thai lexical tone contrasts in monosyllables

While the labels M, H, L, F, and R are used to describe these tones, acoustic analyses have indicated that the actual phonetic shapes of the individual tones – even in citation form – do not always match the labels well (see, e.g., Abramson, 1962; Erickson, 1974; and Zsiga and Nitisaroj, 2007). None of the five tones are actually completely level. The M tone is closest to level, as it stays within the middle of the pitch range, but even it falls approximately 20 Hz across a syllable (Zsiga and Nitisaroj, 2007). The H tone is a scooped contour: it falls slightly and remains as low as (or even lower than) the M tone for the first half of the syllable. It then rises steeply in the second half of the syllable. The L tone falls steadily across a syllable and reaches the bottom of the pitch range at the syllable's end. The F tone has a rise-fall contour, and the R tone has a fall-rise contour (Zsiga and Nitisaroj, 2007).

According to Zsiga and Nitisaroj (2007), no consensus has yet been reached on the identity of the TBU in Thai. Suggestions have included the vowel (Gandour, 1974; Leben, 1971, 1973); the syllable (Yip, 1982; Zhang, 2002; Yip, 2002); and the mora or syllable (Yip, 2002). In addition, Morén and Zsiga (2006) and Zsiga and Nitisaroj (2007) posit a moraic alignment hypothesis which, in short, suggests that Thai H and L pitch targets are aligned to morae. For the purposes of this study, I assume the TBU to be the vowel.

In addition, the same issue over whether Mandarin contour tones are unitary or compositional is debated about Thai contour tones. Zsiga and Nitisaroj (2007) note that it is possible to compose complex Thai contours from simple H and L levels borne by a syllable, but that such phonetic mapping rules would be complex. A single H borne by a syllable would need to be mapped into a level-rising scooped contour, while an H linked as part of a falling tone would be mapped to a very rapid rise to the top of the pitch range. Such complexity leads Abramson (1979:7) to reject a compositional analysis of the contours, arguing that the data lend no phonetic plausibility to arguments for the specification of R and F tones as compositional sequences of H and L tones. Here again, for this study, I adopt the unitary view.

Abramson (1962, 1975) reports that native listeners can identify each tone on the basis of F0 alone: when five synthetic average F0 contours were imposed onto syllables (creating a set of tonal minimal quintuplets), native listeners' identifications were near or at ceiling. On the other hand, whispered (toneless) syllables are not well identified (Abramson, 1972). It therefore appears that F0 cues are more salient than other acoustic cues (e.g., duration and amplitude) for native-listener perception. More specifically, F0 direction may be of greater importance than offglide F0 for tone perception (Pike, 1948; Gandour, 1983). Abramson (1978) reported that contour-tone slope is also important for Thai tone perception. Level tone trajectories were usually identified as H, M, or L tones, but the addition of pitch movement over the syllable aided perception. For instance, that which was most reliably identified as an H began at the middle of the pitch range and rose 30 Hz across the syllable, while that which was most reliably identified as an L began in the middle of the range and fell 30 Hz across the syllable. It seems, therefore, that the tonal contrasts of Thai are defined in terms of pitch change direction and slope and direction of pitch change. Finally, the timing of pitch inflections may be essential cue for Thai tone perception (Gussenhoven, 2004; Shen and Lin 1991; Xu 1998, 1999a, 1999b; and others). H tones are sometimes produced with a final fall, in which case some talkers produce both H and F tones with rise-fall contours (Gandour et al., 1991). Gandour and colleagues surmise that, for those talkers, the primary difference between the H and F tones is the timing of their respective peaks (the turning point) across the contour: the F tone has an early peak while the H tone has a late peak.

Cantonese

Cantonese is a Sino-Tibetan language with three contour and three level tones; it is spoken throughout regions of China, including Hong Kong. The level tones are H (55), M (33) and L (22); they are differentiated via relative F0 level (H is highest, L is lowest, and M is in the middle). They are similar in that their F0 contours change little across their trajectories. Contour tones differ with regard to the direction and magnitude of F0 change (Khouw and Ciocca, 2006 and others). The MR (25) and LR (23) tones rise, but the latter has a smaller F0 change than the former. The LF (21) tone falls by a relatively small degree. (Khouw and Ciocca, 2006; Bauer and Benedict, 1997; Fok Chan, 1974; Wong and Diehl, 2003; Francis et al., 2008; and others). Table 2.5, from Wong and Diehl (2003), displays the Cantonese phonemic lexical tone contrast.

Syllable	**Tone**	**English gloss**
si	H	teacher
si	MR	history
si	M	to try
si	LR	market
si	L	yes
si	LF	time

Table 2.5. Cantonese lexical tone contrasts in monosyllables

The LF tone is often produced with some amount of glottalization, but this property has been shown to not function as a consistent perceptual cue for native Cantonese listeners (Vance, 1976). F0 is thought to be the primary – and possibly the sole – acoustic cue for Cantonese tone perception (Francis et al., 2008; Ciocca, Francis, Aisha, and Wong, 2002; Lee, van Hasselt, Chiu, and Cheung, 2002). Specifically, listeners rely on relative F0 level, direction of F0 change, and magnitude of F0 change for Cantonese tone perception (Fok Chan, 1974; Gandour, 1981, 1983; Vance, 1977). Pitch level in particular has been suggested to be perceptually more salient than pitch contour (Gandour, 1983).

2.3. Data from the current study on the tones of the five languages

The tone data collected in this study appear to largely corroborate the observations made in the literature. Figures 2.1 through 2.5, below, illustrate the tone contours of each of the five languages. These figures summarize data collected in this study. Each figure shows the overall mean F0 in semitones (ST) of the language's citation-form tones across their trajectories, as produced in CV syllables by (and aggregated over) 3 male and 3 female native speakers. F0 outliers, defined as F0 values more than 2.5 Standard Deviations from the mean for that tone for each individual talker, are omitted from these and all subsequent graphs and analyses. That said, only a very small number of the syllables were outliers. In total, 2880 of the 115,209 syllables, or 0.025%, were outliers. To break it down by language, 193 of the 11,609 Igbo syllables (0.02%) were outliers, as were 310 of the 17,041 Yoruba syllables (0.02%); 416 of the 23,228 Mandarin syllables (0.02%); 814 of the 29,087 Thai syllables (0.03%); and 1130 of the 34,244 Cantonese syllables (0.03%). Standard deviation was calculated via the equation

$$\sqrt{\frac{\sum (x-\bar{x})^2}{(n-1)}}\ .$$

As mentioned in chapter one, the semitone scale, a logarithmic transformation of the physical Hertz scale, is used throughout this study. It captures speakers' intuitions about the equivalence of intonational spans (personal pitch ranges), and takes into account one of the primary assumptions of the TAD, that speech-sounds are produced (organized in acoustic space) in such a way as to make them sufficiently distinct for the listener. This is one of several psychoacoustic scales, including mel, Bark, and Equivalent Rectangular Bandwidth (ERB)-rate. The mel scale was used by Liljencrants and Lindblom (1972), but I use the semitone scale because, compared to the mel, ERB-rate, and Bark scales, it more accurately reflect listeners' intuitions about intonational equivalence (i.e., intonational span) (Nolan, 2007). Tonal F0 was measured across the vowel (the presumed TBU) of each CV syllable at ten equidistant timepoints *k* in Hz. Hz measurements were then converted to ST using the equation F0semitones = (12 log(F0Hz/100 Hz))/log(2) (http://www.linguistics.ucla.edu/faciliti/facilities/acoustic/pitch_unit_conversion.txt) via Perl script. Timepoints k1-k9 are shown; the F0 of the initial vocalic segment, at timepoint k0, was excluded because it is perturbed by the preceding consonant (Hombert et al., 1979). Further details about the materials and the methods employed for eliciting the syllables are provided in chapter three.

The figures are ordered according to the number of tones in the languages' inventories. Figure 2.1 illustrates the tones of Cantonese (6 tones); Figure 2.2 illustrates the tones of Thai (5 tones); Figure 2.3 illustrates the tones of Mandarin (4 tones); Figure 2.4 illustrates the tones of Yoruba (3 tones); and Figure 2.5 illustrates the tones of Igbo (2 tones). Beneath each figure is a description of the language's observed tonal trajectories and a discussion of how they compare with the description of that language's tones reported in the literature.

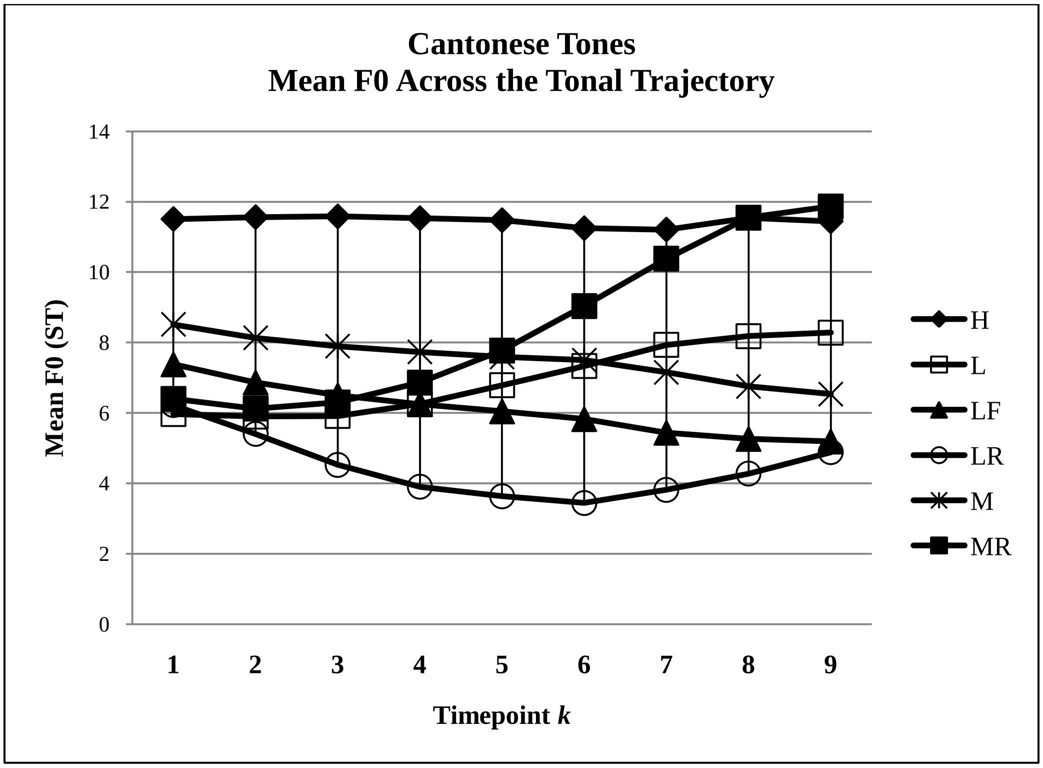


Figure 2.1. Cantonese tonal trajectories in mean F0 (ST)

	Timepoint k								
Tone	**k1**	**k2**	**k3**	**k4**	**k5**	**k6**	**k7**	**k8**	**k9**
H	11.509	11.563	11.583	11.532	11.476	11.246	11.208	11.539	11.445
L	5.965	5.897	5.914	6.257	6.784	7.333	7.932	8.181	8.283
LF	7.381	6.864	6.506	6.247	6.046	5.826	5.439	5.264	5.189
LR	6.219	5.401	4.524	3.901	3.632	3.442	3.815	4.273	4.886
M	8.515	8.131	7.894	7.733	7.589	7.498	7.155	6.753	6.536
MR	6.397	6.120	6.299	6.861	7.773	9.038	10.378	11.547	11.868

Table 2.6. Mean F0 (ST) values of Cantonese tonal trajectories

In keeping with descriptions in the literature, my Cantonese talkers' H tone is approximately level throughout its trajectory, with an average F0 of 11.45 ST. Also as described in the literature, my talkers' LF tone's fall is steady and small in magnitude; it only falls about 2.2 ST.

Also, their MR tone rises sharply (about 5.8 ST). Furthermore, my talkers' LR tone does indeed rise (about 1.5 ST, less sharply than their MR tone), but unlike descriptions in the literature, it only does so after it has fallen 2.8 ST from timepoint k1 to k6. The most significant divergence between my talkers' tone productions and those described in the literature is the acoustic realization of the M and L tones. While the literature suggests the M and L tones are relatively flat, my talkers' M tone steadily falls approximately 2 ST across its trajectory, while their L tone rises about 2.3 ST. The M and L tones' F0 contours are equivalent at timepoint k6, which would presumably be confusing for listeners. However, such potential confusion might be mitigated by attendance to overall pitch-height differences (Gandour, 1983), since the M tone F0 contour is, overall, higher and more level than the L tone F0 contour. Differences between my data and that reported in the literature could be caused by differences between tone-elictation methods across studies; it is more difficult to label tones in citation form than those in sentence context (Wong and Diehl, 2003), and it is likely more difficult to produce tones in citation form than in context.

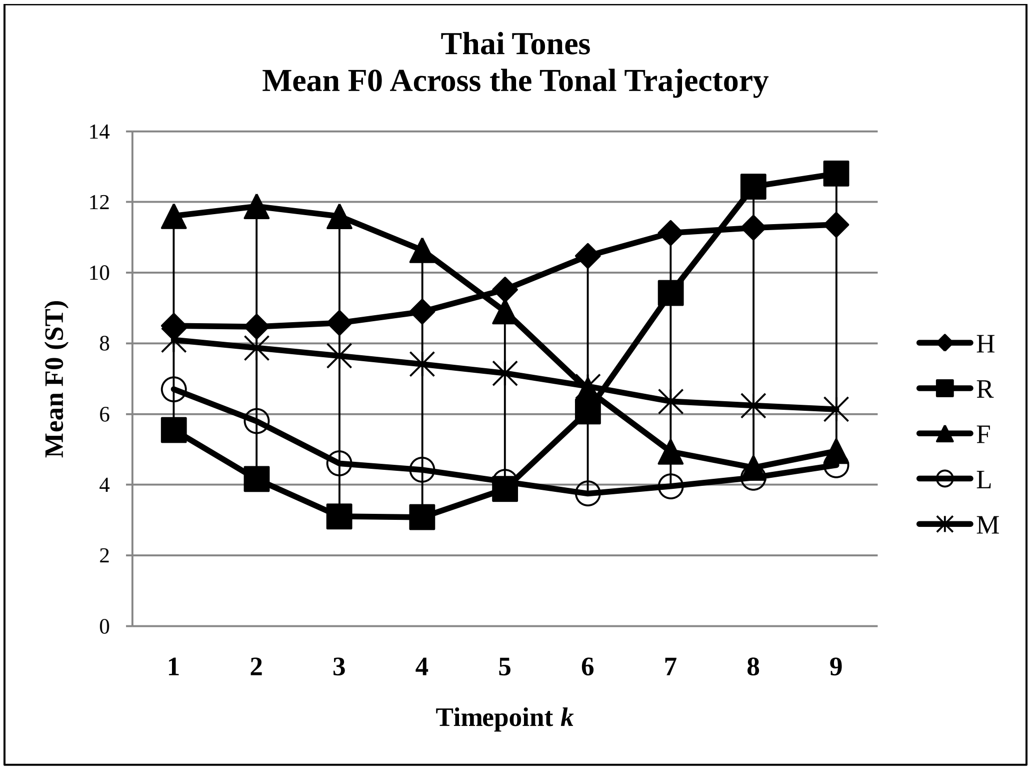


Figure 2.2. Thai tonal trajectories in mean F0 (ST)

	Timepoint k								
Tone	**k1**	**k2**	**k3**	**k4**	**k5**	**k6**	**k7**	**k8**	**k9**
H	8.496	8.474	8.579	8.8986	9.520	10.474	11.124	11.274	11.357
R	5.549	4.161	3.106	3.0832	3.880	6.077	9.424	12.438	12.804
F	11.601	11.881	11.598	10.639	8.905	6.684	4.936	4.480	4.955
L	6.700	5.800	4.605	4.422	4.083	3.754	3.958	4.203	4.552
M	8.098	7.869	7.647	7.4134	7.152	6.782	6.357	6.238	6.137

Table 2.2. Mean F0 (ST) values of Thai tonal trajectories

As described in the literature, my Thai talkers' level tones were not actually level. The M tone was indeed closest to level, and stayed in the middle of the pitch range, but it falls about 2 ST across its trajectory. However, my talkers' H tone differs from that described in the literature: instead of being a scooped contour, it is level from tonal onset to timepoint k3, rises about 2.5 ST until timepoint k7, and levels out to the offglide. Additionally, instead of falling steadily across the syllable and reaching the bottom of the pitch range at offglide, my talkers' L tone falls sharply – about 2.1 ST – from onset to timepoint k3; continues to fall, but to a lesser degree (about 0.8 ST), until timepoint 6; and rises 0.8 ST to the offglide. As described in the literature, my talkers' R tone does have a falling-rising contour: it falls sharply (nearly 2.5 ST) from onset to timepoint k3; flattens out to timepoint k4; and rises 9.7 ST to the offglide. Likewise, their F tone has a rise-fall contour: it rises slightly from onset to timepoint k2, falls gently to timepoint k4; falls steeply from timepoints k4 to k8, and rises slightly to offglide. The magnitude of pitch change for the R and F tones are the greatest of all the tones in the inventory.

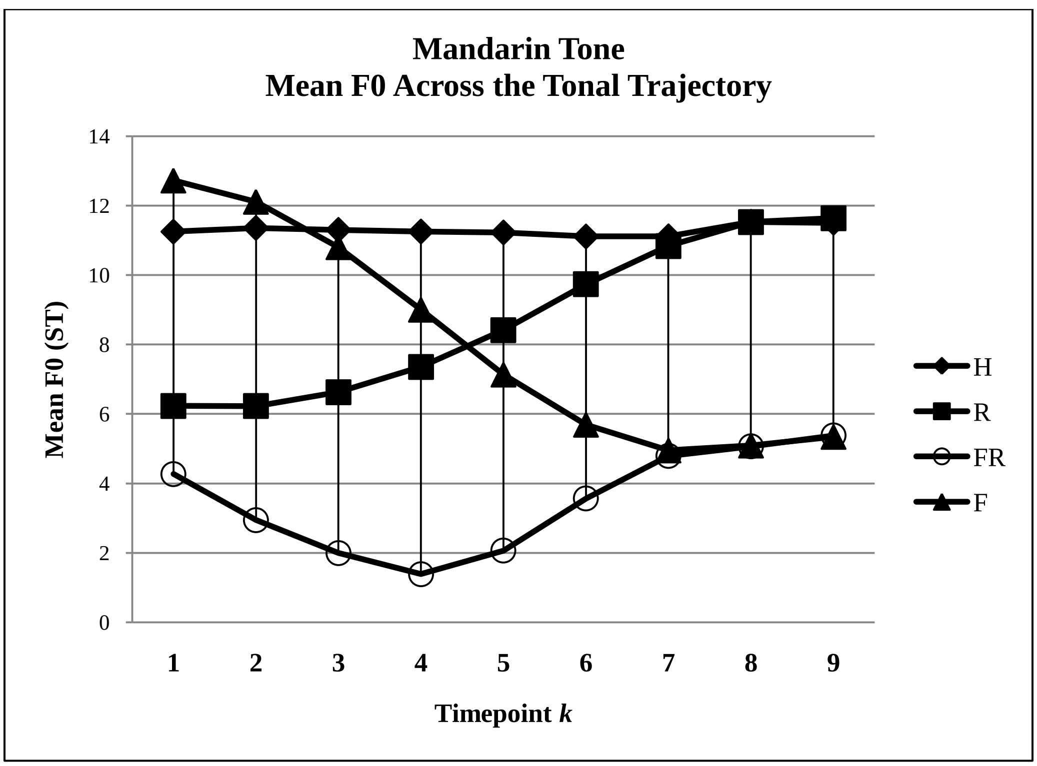


Figure 2.3. Mandarin tonal trajectories in mean F0 (ST)

	Timepoint k								
Tone	**k1**	**k2**	**k3**	**k4**	**k5**	**k6**	**k7**	**k8**	**k9**
H	11.249	11.356	11.301	11.253	11.226	11.111	11.115	11.533	11.505
R	6.230	6.225	6.626	7.356	8.412	9.738	10.838	11.524	11.636
FR	4.270	2.945	1.994	1.386	2.063	3.564	4.792	5.068	5.379
F	12.725	12.112	10.805	9.004	7.134	5.690	4.950	5.090	5.341

Table 2.8. Mean F0 (ST) values of Mandarin tonal trajectories

As described in the literature, my Mandarin talkers' H tone is essentially level. Their R tone starts to rise at about timepoint k2, and actually remains level (does not fall) until that point. It ultimately rises about 5.5 ST total. My talkers' FR tone falls about 2.9 ST, but only until timepoint k4 – its turning point occurs slightly earlier than described in the literature. It then rises sharply (about 3.4 ST) until timepoint k7, and then continues to rise gently to offglide. Finally, my talkers' F tone does not rise in the first 20% of the vowel; it rather falls steadily, by about 7.8 ST, until about timepoint k7 and then flattens out until the offglide.

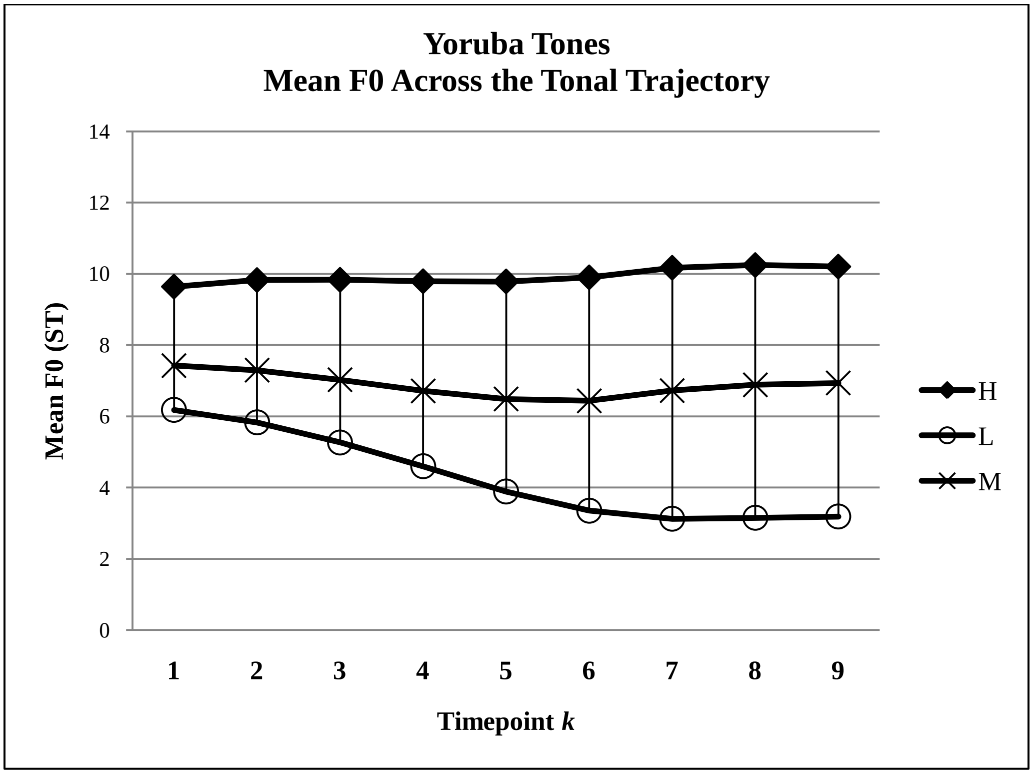


Figure 2.4. Yoruba tonal trajectories in mean F0 (ST)

	Timepoint k								
Tone	**k1**	**k2**	**k3**	**k4**	**k5**	**k6**	**k7**	**k8**	**k9**
H	9.638	9.824	9.832	9.791	9.785	9.901	10.171	10.247	10.203
L	6.182	5.830	5.267	4.598	3.887	3.352	3.125	3.151	3.189
M	7.427	7.295	7.024	6.715	6.487	6.437	6.725	6.887	6.934

Table 2.9. Mean F0 (ST) values of Yoruba tonal trajectories

As suggested in the literature, my Yoruba talkers' H and M tones were approximately level throughout its trajectory, though the M tone dipped slightly around timepoint k6. Likewise, following Hombert (1976)'s observations, the onset of my talkers' L tone is lower in pitch than that of the H and M tones and displays a falling pitch contour. In fact, my talkers' L tone falls about 3 ST until about timepoint k7, but then it remains approximately level until offglide.

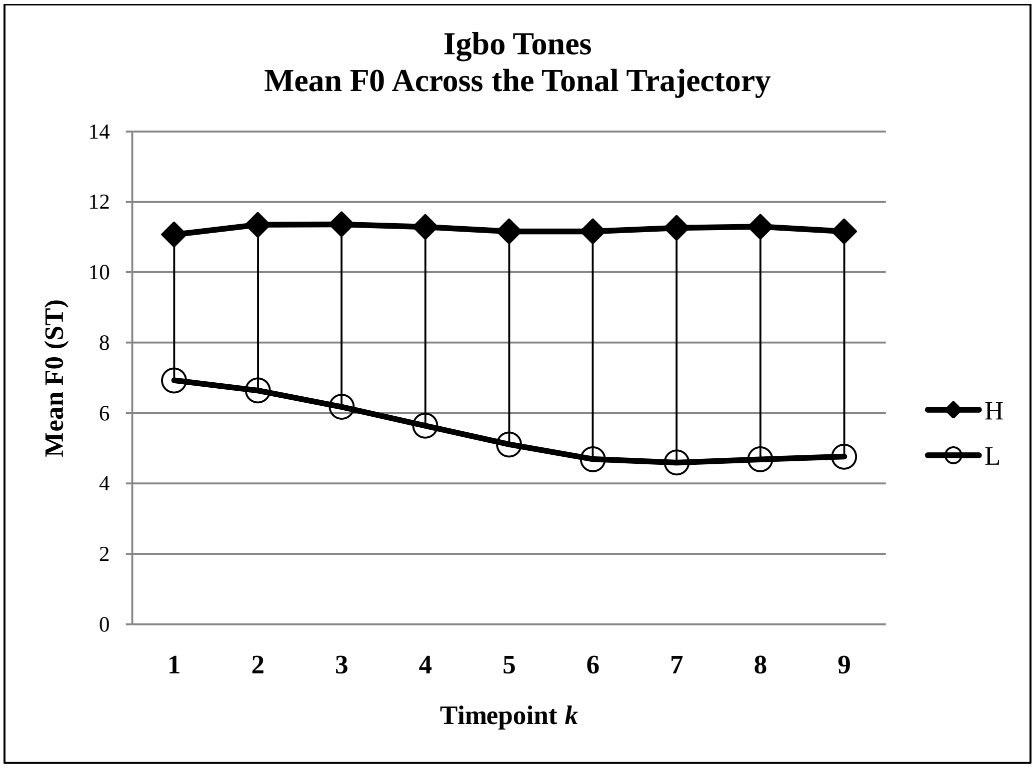


Figure 2.5. Igbo tonal trajectories in mean F0 (ST)

	Timepoint k								
Tone	**k1**	**k2**	**k3**	**k4**	**k5**	**k6**	**k7**	**k8**	**k9**
H	11.071	11.345	11.363	11.287	11.161	11.162	11.259	11.291	11.160
L	6.926	6.640	6.177	5.639	5.107	4.688	4.592	4.682	4.760

Table 2.10. Mean F0 (ST) values of Igbo tonal trajectories

My Igbo talkers' tones were similar in certain respects to Igbo-tone observations in the literature. Their H tone was level throughout its trajectory, but its highest point was found at timepoint k3, not at the end of the vocalic timespan as reported in the literature. Considering how little their H tone changed throughout the trajectory, H tone perception might not actually be impacted. My talkers' L tone steadily fell about 2.3 ST from onset to timepoint 7, then rose slightly to offglide.

The L tone target was also not found at offglide – it instead was found at timepoint k7 – but the tonal F0 at timepoint k7 was only 0.17 ST lower than that at offglide. L tone perception, too, might be unaffected by this small a difference. Regardless, the H and L tones were highly differentiated (separated by 6.4 ST) at offglide.

CHAPTER THREE: DATABASE COMPILATION

3.1. Introduction

This chapter is organized in the following manner: in section 3.2., I describe the materials collected and used for this study, and in section 3.3., I describe the methods employed for their collection. In brief, the participants in this study, 3 male and 3 female native speakers of each language, were recorded as they produced 18 CV syllables with each contrastive tone. The start and end of the tone-bearing unit (TBU) of each syllable – its vowel – was delineated, and the vocalic F0 (in Hz) was measured at 10 equidistant timepoints *k*. The Hz values were converted to semitones (ST); these ST values were used for the analyses described in chapters four and five of this dissertation.

3.2. Materials

The tones analyzed in this study were borne by several isolated CV syllables, namely, [ba], [bi], [bu], [da], [di], [du], [ga], [gi], [gu], [la], [li], [lu], [ma], [mi], [mu], [na], [ni], and [nu]. The consonants [b], [d], [g], [l], [m], and [n] were chosen for multiple reasons. They, or consonants that are comparable in place of articulation, are used across the five languages in initial position. (While Cantonese and Mandarin lack voiced plosives, they have [p] and [p^h], [t] and [t^h], and [k] and [k^h]; the former (unaspirated) of each pair is similar to [b], [d], and [g] in place of articulation and occur in initial position. Similarly, Thai lacks [g] initials, but has [k] and [k^h], the former of which is similar in place of articulation and voicing to [g].) Choosing these consonants in particular minimized the possibility that participants (talkers) would be confused when asked to produce them. Voiced consonants were chosen for two main reasons:

(1) because both they and the vowel are voiced oral segments, voicing and oral airflow will be uninterrupted throughout the duration of the syllable; and (2) because they are obstruents, it will be relatively easy to identify the vocalic onset. Similarly, the vowels [i], [a], and [u] were chosen because the five languages all have them in their inventories, and all use them in coda position.

Not all the syllables are real (meaningful) words in all five languages, as it is not possible to find a complete set of phonetically-identical CV syllables that are all meaningful real words in all the languages. Native-speaker language consultants (one per language, for a total of five) were hired to identify which syllables were real words and which were non-words. They provided glosses for each of the real words, and they wrote short sentences in the native language orthography (plus their English glosses) to exemplify each real word in context. The Mandarin and Yoruba language consultants also translated, from English to their native languages, a passage called The North Wind and the Sun (English version from International Phonetic Association, 1999), to be used in a future project. Finally, language consultants translated, from English to their native language, the phrase *Do you speak [language]?* for use on recruitment materials. *Appendix A: Materials* contains the lists of syllables and passages. Each real word is listed with its gloss, and each non-word is marked with dashes. Language consultants were paid at the rate of $10 per hour. Funds were provided by a Northwestern University Graduate Research Grant to Jennifer A. Alexander and by the Department of Linguistics at Northwestern University.

3.3. Methods

3.3.1. Participants

Three female and three male adult native speakers of each language produced the syllables. As all the talkers lived in the U.S. at the time of testing, all spoke and understood English to some degree, but all listed English as a non-native language. All were literate in both English and their native language, and none reported any speech or hearing problems. Place-of-origin was controlled to the extent possible in order to minimize dialect variation across talkers in each language group. Igbo and Yoruba speakers had spent the majority of their lives in Nigeria; Mandarin speakers in Beijing, China; Thai speakers in Bangkok, Thailand; and Cantonese speakers in Hong Kong, China. Save for five exceptions, all participants had lived in his/her place-of-origin for at least their first 13 years of life; most had lived in his/her place of origin considerably longer (average age at which participants left their place-of-origin = 21.3 years of age). The exceptions included one Igbo speaker and two Yoruba talkers whose information was unreported; one Cantonese participant who moved to Hong Kong from Guangdong Province, China, at age 5 and lived there until the age of 29; and one Cantonese participant who moved to Hong Kong from Pittsburgh, PA, US at age 2 and lived there until the age of 27. Information about the participants' ages at time of testing and place(s) of residence until the point at which they immigrated to the U.S. is provided in *Appendix B: Participants*.

Participants were recruited and run between February and August 2009. They were recruited via IRB-approved flyers, emails, and Craigslist (online) ads. Most of the Cantonese, Thai, and Mandarin talkers were recruited at Northwestern University. They contacted the study coordinator (me) at the email address provided (a) on recruitment flyers and (b) within the text of emails forwarded to international-student listservs from the leaders of those organizations. The majority of the Thai participants, in fact, learned about the study from the head of the NU Thai Club. Very few Yoruba and Igbo participants responded to ads, so they were primarily recruited from around the Rogers Park and Uptown neighborhoods of Chicago, which both are home to sizeable communities of Nigerian immigrants. In particular, Yoruba immigrants were recruited in person at area churches – after a short presentation about the study, interested individuals had provided their names and contact information. One Igbo participant was recruited via a flyer posted at Northwestern University, but most others were recruited in person at the 2009 Igbo Festival in the Rogers Park neighborhood of Chicago.

Participants were between the ages of 18-50 years (mean = 30.6 years) when they produced the syllables. Due to difficulty incurred in recruiting Nigerian participants that were closely matched in age to the East Asian participants (details discussed below), the East Asian participants were on average about 16 years younger than the Nigerian participants. At the time of testing, participants had resided in the U.S. anywhere from 2 months to 26 years. This range of time spent living the in the U.S. is a result of the age difference between the East Asians and Nigerians – the average length of residence in the U.S. up until the time of testing was approximately 2 years for East Asians but nearly 13 years for Nigerians (excluding 3 Nigerians whose date of immigration was unreported). These details are also included in *Appendix B.*

3.3.2. Recording procedures

Upon arrival, participants signed IRB-approved consent forms and filled out a questionnaire about their language background and all the towns in which they had lived.

Information and instructions about the task were presented in writing on the computer monitor. They were written in both the participant's native language and in English, and were also read aloud, in English, by the experimenter. To continue to subsequent pages of information and instructions, the experimenter verbally checked for comprehension, and the participant pressed the space bar. As a sample, the Cantonese instructions are in *Appendix C: Instructions*.

Syllables were presented one at a time, via Dell Inspiron 600m notebook PC and E-Prime (Psychology Software Tools). Each syllable was presented in the language's orthography; in Roman letters; with tone numbers, letters, and/or diacritics; and with the example sentence written by the language consultants (real words only). An example of a Mandarin trial is shown Figure 3.1, below.

Chinese character	*PinYin*	*Tone #*	*Meaning (in English)*	*Example*	*Example (in English)*
八	bā	ba1	eight/8	我有八个本子。	I have eight notebooks.

Figure 3.1. Mandarin trial

Each syllable, with each contrastive tone, was produced in isolation (to ensure consistent standard pronunciation, and to avoid list intonation when reading the syllables). Participants were instructed to read each one aloud, just once, concentrating on its tone. They were permitted to repeat any syllable if they decided they were dissatisfied with that utterance. The procedure was self-paced; participants could take as long as desired to think about each syllable before they produced it. They then pressed the space bar to continue to the following syllable. Before starting the test trials, participants performed several practice trials (identical to test trials, but with non-test syllables), and were given the option to repeat the practice trials as many times as desired before continuing to the test trials. Most chose not to repeat the practice trials. Participants were provided bottles of water to drink so as to minimize vocal fatigue.

Test trials were organized into six blocks: three blocks in which the syllables were randomized, and three in which the syllables were ordered sequentially. No syllable was ever presented more than once within a block. Thus, each syllable was produced six times by each talker. Participants were either presented with all three sequential-order blocks first and all three random blocks second, or vice-versa (see *Appendix B* for details on each participant). At the end of each block, the experiment stopped automatically; participants took a two-minute break before continuing on to the next block.

Due to the different number of tones in each language, talkers produced different numbers of syllables depending on their native language. Specifically, each Igbo talker produced 216 syllables (18 syllables x 2 tones x 6 blocks); each Yoruba talker produced 324 syllables total (18 syllables x 3 tones x 6 blocks); each Mandarin talker produced 432 syllables total (18 syllables x 4 tones x 6 blocks); each Thai talker produced 540 syllables total (18 syllables x 5 tones x 6 blocks); and each Cantonese talker produced 648 syllables total (18 syllables x 6 tones x 6 blocks). Thus, a grand total of 12,960 syllables were produced ((216 syllables x 6 Igbo talkers = 1296 Igbo syllables) + (324 syllables x 6 Yoruba talkers = 1944 Yoruba syllables) + (432 syllables x 6 Mandarin talkers = 2592 Mandarin syllables) + (540 syllables x 6 Thai talkers = 3240 Thai syllables) + (648 syllables x 6 Cantonese talkers = 3888 Cantonese syllables)).

Each block of syllables was recorded as one continuous mono (left)-channel WAV file at 44.1 kHz with a Marantz Professional Solid State Recorder, model PMD670, and a Shure WH20XLR Dynamic Headset Microphone. The short burst of sound created when participants pressed the space bar between trials was recorded on the right channel, via an ARTcessories Zdirect Professional Passive Direct Box, for use during delineation of the consonant and vowel of each syllable (explained below). In addition, after all the syllables were recorded, each participant read aloud two passages. The first, *The North Wind and the Sun*, was written and subsequently read in the participants' native language. Igbo, Thai, and Cantonese versions were from The Handbook of the International Phonetic Association (1999); Mandarin and Yoruba versions were as translated by Mandarin and Yoruba language consultants. The second, *The Stella elicitation paragraph* from the Speech Accent Archive (Weinberger, http://accent.gmu.edu) was written and read in English. The passages were not analyzed in this study, but rather were acquired for use in future research projects. Recordings were transferred to a Dell desktop PC via a SanDisk 512 MB compact flash card, modelSDCFB and a SanDisk ImageMate CF reader, model SDDR-92. Most talkers took 60-90 minutes to complete the task. As would be expected, Igbo participants typically finished the task within 60 minutes, as they had comparatively few syllables to produce. Cantonese talkers, on the other hand, took about 90 minutes to complete the task, as they had a comparatively high number of syllables to produce. Like the language consultants, participants were compensated for their time at a rate of $10/hour; again, funds were provided by the aforementioned Northwestern University Graduate Research Grant and the Northwestern University Department of Linguistics.

Depending on their schedule and preference, participants performed the task in a quiet room in one of five locations: (1) a phonetics/phonology laboratory at Northwestern University (in sound-attenuated booths); (2) the Edgewater branch of the Chicago Public Library; (3) the Uptown branch of the Chicago Public Library; (4) a church in the Uptown neighborhood of Chicago; or (5) their private residence (this occurred just once). The vast majority of the East Asian participants performed the task in location (1), and most of the Nigerian participants performed it at locations (2) and (3). No matter the location, the equipment – being portable – was the same, so as to minimize differences in recordings.

3.3.3. Data processing

Data were processed, and analyses conducted, with a Macintosh OSX, 2GHz Intel Core 2 Duo iMac and a Dell Inspiron 1420 notebook PC. Only vowels were analyzed, as the vocalic segment was more consistently modally voiced – and was therefore more conducive to F0 analysis – than the preceding consonant. To organize each recording, a short burst of sound was inserted after each syllable. This burst of sound had been recorded onto the right channel when participants pressed the space bar between trials. The program used for this step was Triggerwave (Chan, 2009, http://groups.linguistics.northwestern.edu/documentation/triggerwave_home.html). The Penn Phonetics Lab Forced Aligner (Yuan and Liberman, 2009), via HTK HVITE (Young, Evermann, Gales, Hain, Kershaw, Liu,… and Woodland, 2006), was then used for transcription of the syllables. HTK HVITE is a forced aligner designed to create transcriptions of recordings at the word level; the Penn Aligner adds a python script that directs HTK to transcribe recordings at the phonetic level. Additionally, a Perl script (Chan 2010, http://groups.linguistics.northwestern.edu/documentation/nualigner_home.html) wrapped around the Penn Aligner made possible batch-processing of audio files, which was essential for my project, as each recording consisted of all the syllables produced within a block. To transcribe my syllables, the aligner took a recording; a list of the syllables in the order in which they were produced in that particular block; and a dictionary custom-made for each language that contained each syllable, its tone, and a transcription of the segments in ASCII text ([b] = B, [d] = D, [g] = G, [l] = L, [m] = M, [n] = N, [a] = AA, [i] = IY, [u] = UW). The aligner returned a Praat text grid (Boersma, 2010) with two tiers: the full syllable and its tone on the top tier, and the consonant and vowel, written in ASCII script, demarcated on the bottom tier. This was a useful, though coarse, first pass at demarcating vocalic onset and offset. The aligner is somewhat limited in its ability to precisely identify segment boundaries. In particular, segment boundaries were often mis-identified when phonation was non-modal (in these cases, breathy or creaky). The aligner also failed to detect syllables at all if the trigger volume was too low. Therefore, its output required careful hand-correction. Each recording and text grid was opened with Praat, and the start and end of each vowel was carefully and consistently demarcated by hand. Specifically, the start of the vowel was measured as the start of vocalic modality, i.e., at the first glottal pulse of the first repeating vocalic wave. The end of the vowel was measured as either the final glottal pulse – in cases where the end of the vowel was modal – or the end of the final wave, when the end of the vowel was non-modal (creaky or breathy).

F0 (in Hz) of each vowel at ten equidistant points k ($k_{0\text{-}9}$) (following Zhao and Jurafsky 2007; 2009) was automatically measured via a Praat script. The script, originally written by

Liennes (2003), was modified to read sound files and TextGrids as input and to use PitchTier to analyze F0 from labeled segments in the text grid files. (The previous version used PitchObject, which was in many cases unable to measure F0 from the text grid; PitchTier was much more successful.) Pitch minima and maxima were set at 25 and 600 Hz, respectively. This range accommodated variation in the talkers' pitch ranges, with room to spare to make sure that no exceptionally high or low utterances were missed. The Praat script returned a pitchresults text file with the F0 (in Hz) of each vowel at ten equidistant points.

The Hz measurements were then converted to semitones (ST), a logarithmic transformation of the physical Hz scale that, compared to other psychoacoustic scales such as mel, Bark, and ERB-rate, most accurately reflects listeners' intuitions on intonational equivalence (Nolan, 2007). A Perl script read the pitchresults text file and converted Hz to ST using the conversion equation $F0_{semitones} = (12 \log(F0_{Hz}/100 \text{ Hz}))/\log(2)$ (http://www.linguistics.ucla.edu/faciliti/facilities/acoustic/pitch_unit_conversion.txt). The basis of this equation is the musical semitone scale, where each octave equals 12 semitones. The steps of the ST scale thus correspond to equal perceptual intervals; it captures a key psychoacoustic assumption of the TAD and of this study, that talkers intentionally produce tones so as to make them maximally distinct for the listener. The Perl script returned text files that listed information about each vowel, including the F0 at ten equidistant points in both Hz (as originally measured) and in ST. The vocalic F0 measurements, in ST, were used for all analyses in this study.

3.4. Database

On the consent forms, participants were asked to give or deny permission for their recordings to be made available to the general public. Recordings and associated materials (Praat text grids; lists of syllables; demographic information, etc.) for which sharing permission was granted will be uploaded to OSCAAR. These password-protected files will be organized into a searchable database. After a user obtains permission for use, he/she will be able to access any of the files he/she requires. In keeping with IRB regulations, at no time will the participants' names be associated with any of his/her downloadable files. Each participant will be identified by laboratory code only.

CHAPTER FOUR: CROSS-LANGUAGE COMPARISONS AND THE THEORY OF ADAPTIVE DISPERSION

4.1. Introduction

Recall that my overall objective is to illuminate cross-linguistic tendencies in tone system organization. I do so by extending the Theory of Adaptive Dispersion to tone systems. To that end, this study analyzes and compares the sizes of the acoustic lexical-tone spaces, and dispersion of the tones within those spaces, of five languages with very different tone-inventory compositions: Cantonese (6 tones [3 contour, 3 level]), Thai (5 tones [2 contour, 3 level]), Mandarin ([4 tones [3 contour, 1 level]), Yoruba (3 tones [0 contour, 3 level]), and Igbo (2 tones [0 contour, 2 level]).

I follow Zhao and Jurafsky (2007, 2009) and examine one acoustic correlate across the languages: mean overall F0 at various points along the tonal trajectory. I define *tone-space size* as the tonal pitch range, averaged across talkers, measured at three points along the tonal trajectory (tonal onset, midpoint, and offglide). That is, the size of the acoustic tone space is measured as the F0 difference in semitones (ST) between the mean F0 of a language's highest (top) tone and the mean F0 of its lowest (bottom) tone at those three timepoints. I define the comparative *degree of tonal dispersion* as the cross-language difference in the Euclidean distance from the mean F0 (ST) of a given tone relative to the mean F0 (ST) of a tonal *baseline* (namely, the H tone). (This H-tone baseline, a common point of comparison, is so chosen because it is both common to, and phonetically similar in, the languages under comparison.) For example, I compare at tonal onset the F0 difference (in ST) between the M tone and H baseline tone of Cantonese to the F0 difference (in ST) between the M tone and H baseline tone of

Yoruba. If the former is larger than the latter, then the Cantonese M tone is considered comparatively further dispersed from the tonal baseline at tonal onset.

Following the TAD, I assume that tone categories will act as repellers in a dynamical system: each will repel the others and will find equilibrium where it is maximally distant from surrounding tone categories. Closely related to this is my TAD-based assumption that tones will be dispersed only and exactly to the degree necessary to ensure sufficient tonal contrast. A third assumption naturally follows: the distance between two adjacent tone categories in a language will equal the distance between two other adjacent tone categories.

I plan to test the following two competing hypotheses and their accompanying predictions:

H1. Tone spaces will be equivalent in size across languages, and degree of tonal dispersion will differ as a function of tone-inventory size.

a. The size of the acoustic tone space is independent of the size of the tone inventory. With regard to the languages under investigation, this leads to the prediction that Cantonese = Thai = Mandarin = Yoruba = Igbo in overall tone-space size.

b. If tone-space size is equivalent across languages, then the degree of tonal dispersion relative to a tonal baseline will be greater in a language with fewer tones than in a language with more tones. With regard to the languages under investigation, this leads to the prediction that Igbo > Yoruba > Mandarin > Thai > Cantonese in degree of tonal dispersion relative to a baseline.

H2. Tone spaces will differ in size as a function of tone-inventory size, and degree of tonal dispersion will be equivalent across languages.

a. The size of the acoustic tone space is positively correlated with tone inventory size. With regard to the languages under investigation, this leads to the prediction that Cantonese > Thai > Mandarin > Yoruba > Igbo in overall tone-space size.

b. If a language with a larger tone inventory has an expanded tone space relative to a language with fewer tones, the degree of tonal dispersion relative to a tonal baseline will be equivalent across languages. With regard to the languages under investigation, this leads to the prediction that Cantonese = Thai = Mandarin = Yoruba = Igbo in degree of tonal dispersion relative to a baseline.

Figures 4.1 and 4.2, partial reproductions of Figures 1.1 and 1.2, are idealized illustrations of the five languages' tone spaces and degree of dispersion of a given tone within the tone space (here, tone 2) relative to a tonal baseline (tone 1). The highest (top) tone for all languages is called tone 1. The lowest (bottom) tone is 6 in Cantonese, 5 in Thai, 4 in Mandarin, 3 in Yoruba, and 2 in Igbo. Tones are indicated in the abstract (with numbers) because, as discussed later, the highest and lowest positions may be occupied by different tones, depending on the timepoint. Note that in Figure 4.1, which corresponds to H1, the size of the overall tone-space is fixed (the same) across the languages, but the degree of dispersion of tone 2 relative to the baseline tone 1 is largest for Igbo (with 2 tones) and smallest for Cantonese (with 6 tones). In Figure 4.2, which corresponds to H2, the size of the overall tone-space is largest for Cantonese and largest for Igbo; the degree of dispersion of tone 2 relative to the baseline tone 1 is equivalent across the languages.

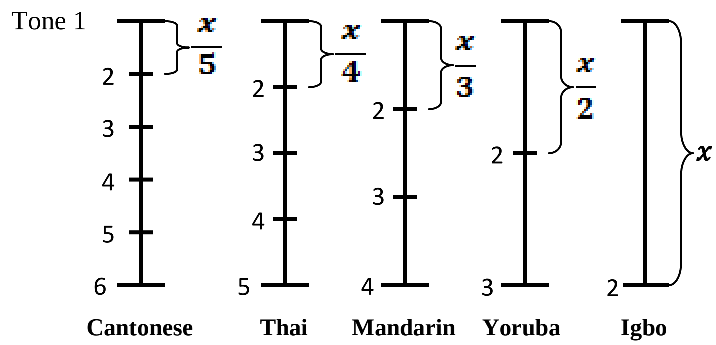


Figure 4.1. An idealized illustration of the five languages' tone-space areas and degree of tonal dispersion under hypothesis H1

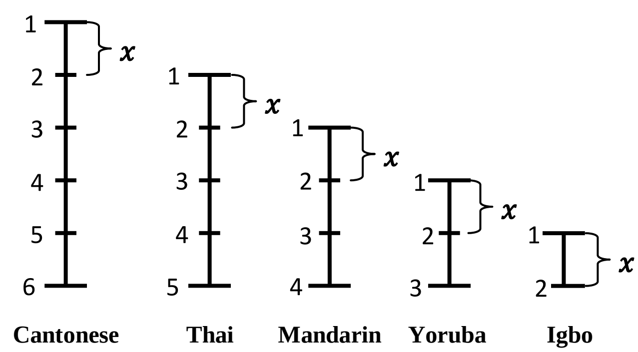


Figure 4.2. An idealized illustration of the five languages' tone-space areas and degree of tonal dispersion under hypothesis H2

The upcoming sections are as follows: In section 4.2, I describe my methods for collecting and analyzing my data. In section 4.3, I examine cross-language tone spaces, and in section 4.4, I examine cross-language tonal dispersion.

4.2. Method

Linear mixed-effects regression models (lmers), fitted using the statistical software package *R* (Free Software Foundation, GNU General Public License) are used to investigate if and how the languages' tone spaces differ from one another as a function of the number of tones in their inventories. Mixed-effects models incorporate two types of factors: fixed (repeatable) and random (non-repeatable, sampled from a larger population); all information about mixed-effects modeling herein is from Baayen (2009). For the analyses, Language and Tone are considered to be fixed variables and Talker and Item are random variables. The Language fixed variable includes any/all of the languages (depending on the model): Cantonese, Igbo, Mandarin, Thai, and/or Yoruba. The Tone fixed variables include the tones under investigation in each model. In the *ToneSpace* analyses, for instance, the tones analyzed are each language's highest and lowest tones at each timepoint *k*. For purposes of analysis, tones within categories being compared are given the same labels, so that *R* can make pairwise comparisons. For example, in the *ToneSpace* analyses, the highest (top) tone was renamed T and the lowest (bottom) B. The Talker random variables are the codes for each individual talker (CF02, CF03, CF04, CM02, CM03, CM04, IF02, IF04, IF05, IM04, IM05, IM07, MF02, MF03, MF05, MM02, MM03, MM04, TF01, TF04, TF05, TM02, TM04, TM05, YF03, YF05, YF07, YM02, YM05, YM06), where 01-07 = talker number, C = Cantonese, F = Female, I = Igbo, M = Mandarin, T = Thai, and Y = Yoruba. Each Item is the Blocking + Repetition + Syllable (i.e., Random blocking; first Repetition; Syllable [bi], resulting in an Item titled *Rand1Bi*). Note that Blocking (Random-Sequential), Repetition (Random order #1-Random order #2, etc.), Consonant ([b]-[d]-[g]-[l]-[m]-[n]), Vowel ([a]-[i]-[u]), Sex (male-female) and Word Status (word-nonword) are collapsed (not defined as separate variables); the effect of each was therefore not tested. In doing so, I choose to focus on variables that are shown in Figures 2.1 through 2.5 to clearly affect tone space (Language and Tone), and ignore potential changes to the tone space caused by one or more of the other variables. That said, future versions of the model may incorporate Vowel as fixed variables, as vowels are known to have intrinsic pitch (see, e.g., Ewan, 1975). Future models may also include Sex, as females are known to display a larger pitch range and vowel space (Diehl, Lindblom, Hoemeke, and Fahey, 1996) and a larger tone space (when defined as the Euclidean distance in F0 of words from a tonal centroid) (Zhao and Jurafsky, 2009).

Along with the obvious benefit of being able to simultaneously model fixed and random effects, mixed-effects models also are potentially more accurate and powerful. Unlike t-tests and ANOVAs, which compare means of aggregated data, a mixed-effects model takes into account all raw data; data loss caused by aggregation is therefore nonexistent. Fixed effects are modeled by means of contrasts (in *ToneSpaces*, T vs. B tone) and random-effect factors are modeled as random variables with a mean of zero and unknown variance. For instance, the talkers in this study may differ with respect to the H tone F0 values. Across the population, the average adjustment required to account for differences in F0 will be zero, but the adjustments required for individual talkers will vary around zero with some standard deviation (an estimating parameter). Treatment coding is such that one level is selected as the default baseline, or reference level; this is represented as an Intercept. By default, factors are ordered alphabetically, and the first is the Intercept. E.g., for the fixed factor Language, Cantonese is the first Intercept, as it is the first in the alphabetized list of languages (Igbo being second, Mandarin being third, etc.). Likewise, Tone Intercepts are determined in alphabetical order as well. Other levels are coded in such a way that their regression weights are the difference between the mean for that level and the mean for the reference (Intercept) level. For instance, the *ToneSpace* analyses determine whether the T-B F0 difference for the Intercept language at a particular timepoint is *different* from that of the other languages.

4.3. Examination of cross-language tone-spaces

Referring back to Figures 2.1-2.5, it is clear that the languages' tone spaces are larger or smaller depending on either the tone or the timepoint along the tonal trajectory. The following three sets of models – *ToneSpaceOnset*, *ToneSpaceMidpoint*, and *ToneSpaceOffglide* – compare at three equidistant points along the tonal trajectory the F0 distances (in ST) between each language's most extreme (top and bottom) tone values.

ToneSpaceOnset evaluates, at the tonal onset (timepoint k1), the following:

(1) Cantonese: The distance between the H and L tones. Referring back to Figure 2.1, H marks the top, and L marks the bottom, F0 values at tonal onset.

(2) Thai: The distance between the F and R tones. Referring back to Figure 2.2, F marks the top, and R marks the bottom, F0 values at tonal onset.

(3) Mandarin: The distance between the F and FR tones. Referring back to Figure 2.3, F marks the top, and FR marks the bottom, F0 values at tonal onset.

(4) Yoruba: The distance between the H and L tones. Referring back to Figure 2.4, H marks the top, and L marks the bottom, F0 values at tonal onset.

(5) Igbo: The distance between the H and L tones, as these are the only tones in Igbo.

ToneSpaceMidpoint evaluates, at the tonal midpoint (timepoint k5), the following:

(1) Cantonese: The distance between the H and LR tones. Figure 2.1 shows that H marks the top, and LR marks the bottom, F0 values at tonal midpoint.

(2) Thai: The distance between the H and R tones. Figure 2.2 indicates that H marks the top, and R marks the bottom, F0 values at tonal midpoint.

(3) Mandarin: The distance between the H and FR tones. Figure 2.3 shows that H marks the top, and FR marks the bottom, F0 values at tonal midpoint.

(4) Yoruba: The distance between the H and L tones. Figure 2.4 shows that H marks the top, and L marks the bottom, F0 values at tonal onset.

(5) Igbo: The distance between the H and L tones, as these are the only tones in Igbo.

ToneSpaceOffglide evaluates, at the tonal offglide (timepoint k9), the following:

(1) Cantonese: The distance between the MR and LR tones. Figure 2.1 shows that MR marks the top, and LR marks the bottom, F0 values at tonal offglide.

(2) Thai: The distance between the R and L tones. Figure 2.2 indicates that R marks the top, and L marks the bottom, F0 values at tonal offglide.

(3) Mandarin: The distance between the R and F tones. Figure 2.3 shows that R marks the top, and F marks the bottom, F0 values at tonal offglide.

(4) Yoruba: The distance between the H and L tones. Figure 2.4 shows that H marks the top, and L marks the bottom, F0 values at tonal offglide.

(5) Igbo: The distance between the H and L tones, as these are the only tones in Igbo.

4.3.1. ***ToneSpaceOnset***

The *ToneSpaceOnset* models compare the F0 difference (in ST) between the top and bottom tones (henceforth called T for "top" and B for "bottom") of the languages at tonal onset (timepoint k1). The first lmer analysis compares Cantonese to each of the other languages. Subsequent regressions examine the remaining parwise comparisons. Figure 4.3, below, shows the top - bottom tone mean F0 at timepoint k1 for the five languages. Standard error bars surround each data point.

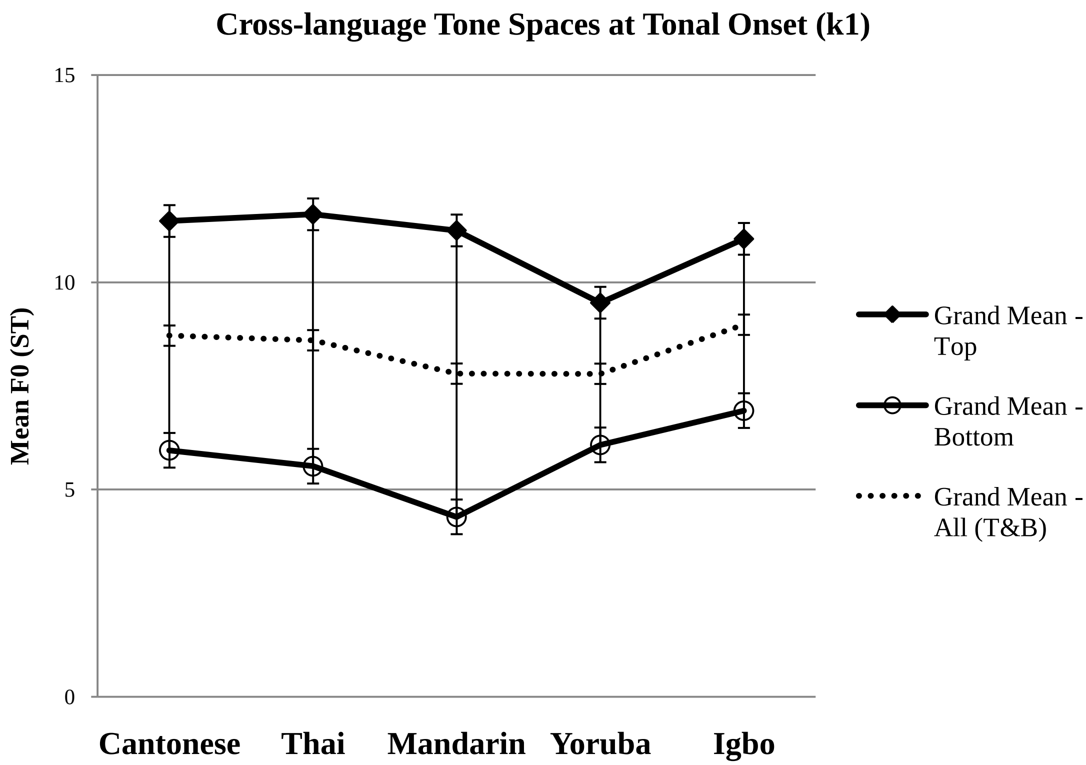


Figure 4.3. Tone-space size across the five languages at the tonal onset

Item	Cantonese	Thai	Mandarin	Yoruba	Igbo
Grand Mean - Top	11.477	11.638	12.772	9.506	11.046
Grand Mean - Bottom	5.947	5.563	4.342	6.078	6.904
Grand Mean – All (T&B)	8.712	8.601	8.557	7.792	8.975
Grand Mean T-B	5.529	6.075	8.431	3.428	4.142

Table 4.1. Tone-space size F0 (ST) values across the five languages at tonal onset

Observe that the Grand Mean – All (T&B) values are very similar across the languages; the lowest Grand Mean value (Yoruba) is only 1.2 ST lower than the highest (Igbo). In addition, observe that the differences between the languages' top vs. bottom tone Grand Mean F0s are all quite different. Table 4.2 shows the fixed-effects results of the *ToneSpaceOnset* lmers. The legend underneath it is to be referenced for this and all other lmer analyses:

ToneSpaceOnset1: Cantonese vs. Igbo, Mandarin, Thai, and Yoruba				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageI	-0.446	2.837	-0.160	0.668
LanguageM	1.278	2.837	0.450	0.2106
LanguageT	0.140	2.837	0.050	0.8832
LanguageY	-2.065	2.837	-0.730	0.0484
ToneB	-5.526	0.123	-44.780	**0.0001**
LanguageI:ToneB	1.379	0.148	9.310	**0.0001**
LanguageM:ToneB	-2.95	0.149	-19.870	**0.0001**
LanguageT:ToneB	-0.57	0.150	-3.810	**0.0006**
LanguageY:ToneB	2.072	0.149	13.920	**0.0001**

ToneSpaceOnset2: Igbo vs. Mandarin, Thai, and Yoruba				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageM	1.724	2.972	0.58	0.105
LanguageT	0.5853	2.972	0.2	0.576
LanguageY	-1.618	2.972	-0.54	0.1306
ToneB	-4.146	0.122	-33.94	**0.0001**
LanguageM:ToneB	-4.333	0.148	-29.3	**0.0001**
LanguageT:ToneB	-1.948	0.149	-13.08	**0.0001**
LanguageY:ToneB	0.6945	0.148	4.69	**0.0001**

ToneSpaceOnset3: Mandarin vs. Thai and Yoruba				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageT	0.396	3.099	0.130	0.2776
LanguageY	-1.814	3.099	-0.590	**0.005**
ToneB	-6.930	0.114	-60.660	**0.0001**
LanguageT:ToneB	0.846	0.141	6.010	**0.0001**
LanguageY:ToneB	3.493	0.140	24.890	**0.0001**

ToneSpaceOnset4: Thai vs. Yoruba				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageY	-2.201	2.838	-0.78	**0.0094**
ToneB	-6.114	0.093	-65.87	**0.0001**
LanguageY:ToneB	2.6333	0.108	24.49	**0.0001**

Legend	
Code	**Gloss**
Est	Estimate
St.E	Standard error
t-val.	t-value
LanguageI	LanguageIgbo
LanguageM	LanguageMandarin
LanguageT	LanguageThai
LanguageY	LanguageYoruba
pMCMC	p-values based on MCMC sampling

Table 4.2. Summary of the results of the *ToneSpaceOnset* lmers

For the fixed effects data above, and all forthcoming analyses, p-values are estimated via Markov chain Monte Carlo (MCMC) sampling, and significant values are in boldface type. The above values are significant at $p \leq 0.0125$ (after Bonferroni correction). The corrected α-level here and in all other analyses was calculated as 0.05/[number of lmers]. The results of the *ToneSpaceOnset* models are summarized below:

1. In general, there is no main effect of language. The Grand Mean F0 did not differ as a function of language, save for the Grand Mean F0 of Yoruba with respect to that of Mandarin (*ToneSpaceOnset3-4*).
2. Overall, there is a significant difference between the top and bottom tones. In each of the models, the bottom tone was 4-7 ST lower on average than the top tone. This indicates that the top and bottom tones are well-differentiated overall.
3. The interaction of tone and language is significant.

Taken together, the results of *ToneSpacesOnset* corroborate the observations of the data in Figure 4.3. Crucially, the models indicate that Yoruba < Igbo < Cantonese < Thai < Mandarin with regard to tone-space size at onset.

The results of the above lmers support neither H1 nor H2: at onset, the languages do not have equivalently-sized tone spaces, and the language with the largest tone inventory (Cantonese) does not have the largest tone space. However, these data suggest that tone-space size at the tonal onset may first be determined by the *type* of tones in the inventory, and then by the *number* of tones in the inventory. Overall, the level-tone-only languages have smaller tone spaces at tonal onset than contour-tone languages. Furthermore, within each language type (level or contour), a smaller tone inventory seems to require a larger tone space. Across level-tone languages, the language with the smaller tone inventory (Igbo) has a larger F0 range than the language with the larger tone inventory (Yoruba). Across contour-tone languages, the language with the fewest tones (Mandarin) has a larger F0 range than languages with more tones; the language with the largest tone inventory (Cantonese) has the smallest F0 space. The tone spaces of level vs. contour tone languages are compared in chapter five.

4.3.2. *ToneSpaceMidpoint*

The *ToneSpaceMidpoint* models compare the F0 difference (in ST) between the top and bottom tones of the languages at tonal midpoint (timepoint k5). The first lmer analysis compares Cantonese to each of the other languages. Subsequent regressions examine the remaining parwise comparisons. Figure 4.4, below, shows the top – bottom tone Mean F0 at timepoint k5 for the five languages. Standard error bars surround each data point.

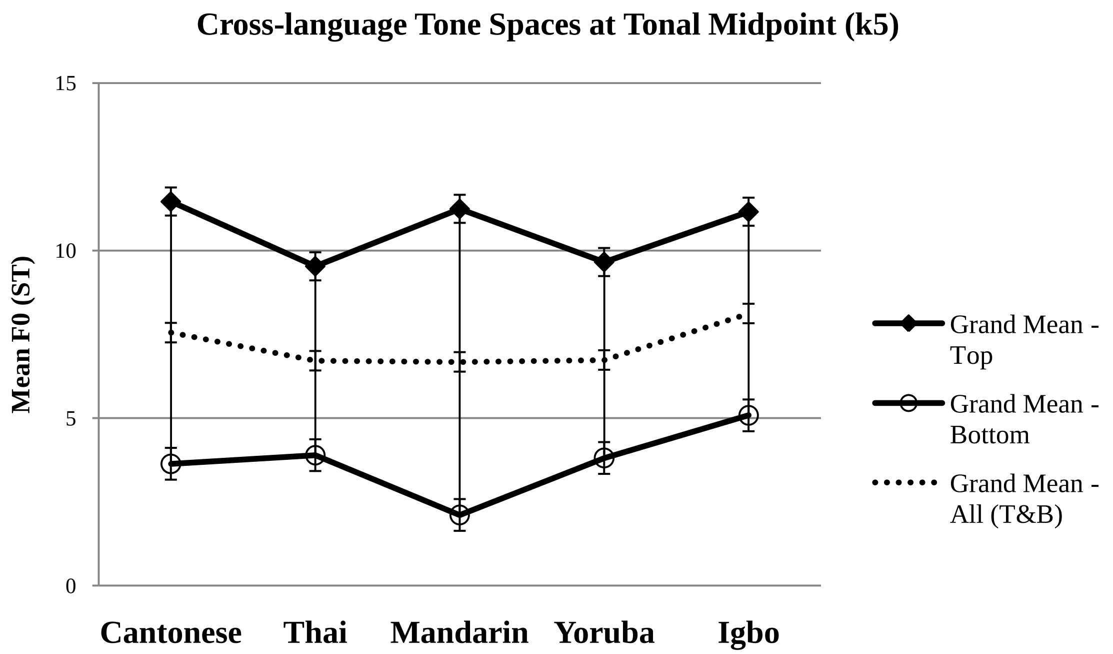


Figure 4.4. Tone-space size across the five languages at the tonal midpoint

Item	Cantonese	Thai	Mandarin	Yoruba	Igbo
Grand Mean - Top	11.466	9.531	11.247	9.659	11.161
Grand Mean - Bottom	3.637	3.895	2.12	3.810	5.082
Grand Mean – All (T&B)	7.552	6.713	6.679	6.735	8.122
Grand Mean – T-B	7.829	5.636	9.137	5.849	6.079

Table 4.3. Tone-space size F0 (ST) values across the five languages at tonal midpoint

Observe that the Grand Mean – All (T&B) values are very similar across the five languages; the lowest Grand Mean (Mandarin) is only 1.4 ST lower than the highest (Igbo). However, the differences between the languages' top vs. bottom tone Grand Means are quite different, save for Yoruba and Thai, which are approximately the same. Table 4.4, below, summarizes the fixed-effects results of the *ToneSpaceMidpoint* lmers.

***ToneSpaceMidpoint1:* Cantonese vs. Igbo, Mandarin, Thai, and Yoruba**				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageI	-0.313	2.606	-0.120	0.8182
LanguageM	-0.245	2.606	-0.090	0.8542
LanguageT	-1.934	2.606	-0.740	0.162
LanguageY	-1.890	2.606	-0.730	0.1568
ToneB	-7.815	0.178	-43.970	**0.0001**
LanguageI:ToneB	1.759	0.222	7.940	**0.0001**
LanguageM:ToneB	-1.325	0.221	-5.990	**0.0001**
LanguageT:ToneB	2.181	0.222	9.810	**0.0001**
LanguageY:ToneB	1.947	0.223	8.750	**0.0001**

***ToneSpaceMidpoint2:* Igbo vs. Mandarin, Thai, and Yoruba**				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageM	0.067	2.747	0.020	0.9608
LanguageT	-1.621	2.747	-0.590	0.2184
LanguageY	-1.577	2.747	-0.570	0.2294
ToneB	-6.056	0.169	-35.840	**0.0001**
LanguageM:ToneB	-3.084	0.202	-15.290	**0.0001**
LanguageT:ToneB	0.420	0.203	2.070	0.0384
LanguageY:ToneB	0.194	0.203	0.950	0.3502

***ToneSpaceMidpoint3:* Mandarin vs. Thai and Yoruba**				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageT	-1.687	2.700	-0.630	0.197
LanguageY	-1.645	2.700	-0.610	0.2092
ToneB	-9.140	0.164	-55.770	**0.0001**
LanguageT:ToneB	3.503	0.202	17.380	**0.0001**
LanguageY:ToneB	3.276	0.202	16.250	**0.0001**

***ToneSpaceMidpoint4:* Thai vs. Yoruba**				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageY	0.040	2.431	0.020	0.9702
ToneB	-5.638	0.147	-38.330	**0.0001**
LanguageY:ToneB	-0.223	0.167	-1.340	0.1898

Table 4.4. Summary of the results of the *ToneSpaceMidpoint* lmers

The above values are significant at $p \leq 0.0125$ (after correction). The results of *ToneSpacesMidpoint* are as follows:

1. <u>There is no main effect of language.</u> In each of the models, the Grand Mean F0 did not differ as a function of language.
2. <u>Overall, there is a significant difference between the top and bottom tones.</u> In each of the models, the bottom tone is 3-9 ST lower on average than the top tone. This indicates that the top and bottom tones are well-differentiated overall.

3. Most of the interactions of tone and language are not significant in *ToneSpaceMidpoint2* and *ToneSpaceMidpoint4*: the Thai and Yoruba tone spaces do not differ in size from that of Igbo, and that the Yoruba tone space does not differ in size from that of Thai.
4. The interaction of tone and language is otherwise significant.

Taken together, the results of *ToneSpaceMidpoint* largely support the observations of the data in Figure 4.4. There was no main effect of language, and the results suggest the following hierarchy with regard to tone-space size at tonal midpoint: Igbo = Yoruba = Thai < Cantonese < Mandarin.

Like *ToneSpaceOnset*, the results of *ToneSpaceMidpoint* support neither H1 nor H2. While two of the languages with larger tone inventories (Cantonese and Mandarin) do have larger tone spaces than the languages with the smallest tone inventories (Igbo and Yoruba), the tone space of Thai was approximately equivalent to those of the languages with the smallest tone-inventory sizes (Yoruba and Igbo). That said, the results of *ToneSpaceMidpoint* may further support the notion that tone-space size is first determined by the type of tones in the inventory, then by the number of tones in the inventory (this is explicitly tested in chapter five). Relative to the contour-tone languages, the level-tone-only languages have smaller tone spaces. Furthermore, across the contour-tone languages, a larger tone inventory has a smaller tone space at tone midpoint. On the other hand, tone-inventory size does not affect level-tone-language tone-space size at midpoint.

4.3.3. *ToneSpaceOffglide*

The *ToneSpaceOffglide* models compare the F0 difference (in ST) between the top and bottom tones of the languages at tonal offglide (timepoint k9). The first lmer analysis compares Cantonese to each of the other languages. Subsequent regressions examine the remaining pairwise comparisons. Figure 4.5, below, shows the top - bottom tone Mean F0 at timepoint k9 for the five languages. Standard error bars surround each data point.

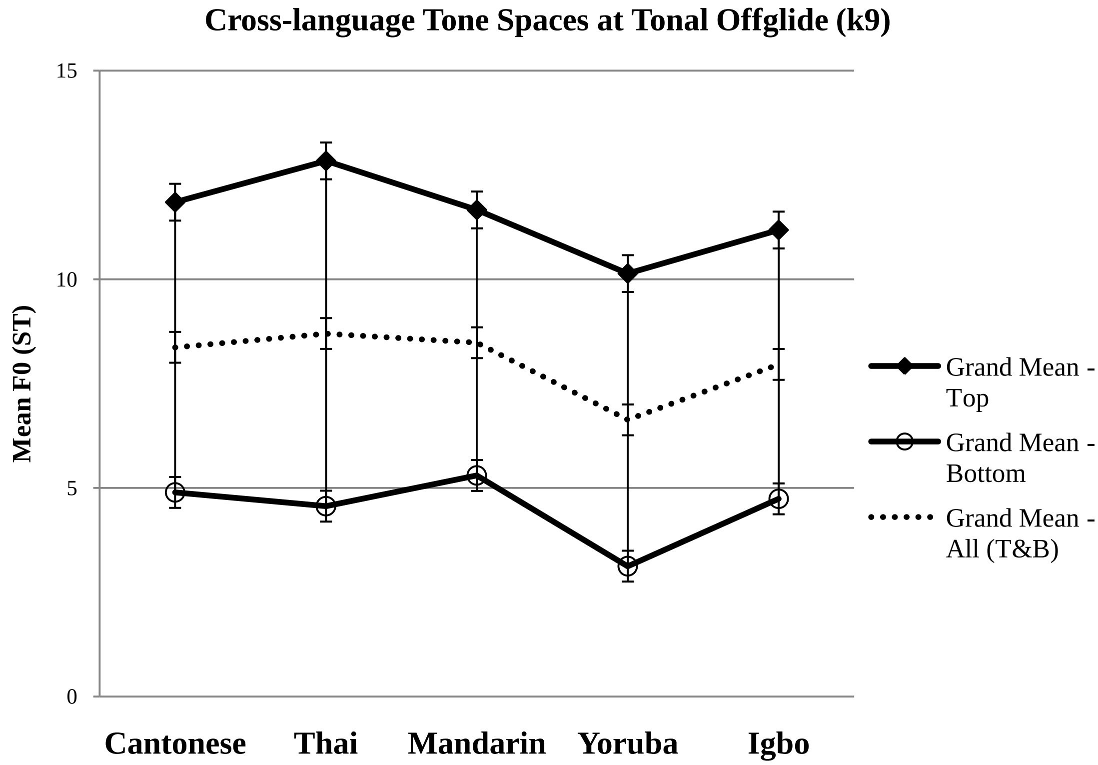


Figure 4.5. Tone-space size across the five languages at the tonal offglide

Item	Cantonese	Thai	Mandarin	Yoruba	Igbo
Grand Mean - Top	11.843	12.834	11.658	10.134	11.177
Grand Mean - Bottom	4.888	4.560	5.295	3.122	4.735
Grand Mean – All (T&B)	8.366	8.697	8.477	6.628	7.956
Grand Mean – T-B	6.955	8.274	6.364	7.011	6.442

Table 4.5. Tone-space size F0 (ST) values across the five languages at tonal offglide

Observe that the Grand Mean – All (T&B) values are nearly the same for Cantonese, Thai, Mandarin, and Igbo, and the lowest Grand Mean (Yoruba) is about 2 ST lower than the highest (Thai). The languages' top vs. bottom tone Grand Mean F0 differences vary, but the difference between Igbo and Mandarin, and the difference between Yoruba and Cantonese, are very small. Table 4.6, below, shows the fixed-effects results of the *ToneSpaceOffglide* lmers.

***ToneSpaceOffglide1:* Cantonese vs. Igbo, Mandarin, Thai, and Yoruba**				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageI	-0.685	2.562	-0.267	0.7178
LanguageM	-0.229	2.562	-0.089	0.9052
LanguageT	0.950	2.562	0.371	0.621
LanguageY	-1.792	2.562	-0.700	0.3484
ToneB	-6.944	0.276	-25.186	**0.0001**
LanguageI:ToneB	0.493	0.387	1.272	0.1912
LanguageM:ToneB	0.579	0.387	1.498	0.1366
LanguageT:ToneB	-1.312	0.387	-3.390	**0.0012**
LanguageY:ToneB	-0.081	0.390	-0.207	0.8402

***ToneSpaceOffglide2:* Igbo vs. Mandarin, Thai, and Yoruba**				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageM	0.460	2.782	0.165	0.8108
LanguageT	1.639	2.782	0.589	0.3684
LanguageY	-1.103	2.782	-0.397	0.5496
ToneB	-6.466	0.242	-26.760	**0.0001**
LanguageM:ToneB	0.078	0.330	0.241	0.8074
LanguageT:ToneB	-1.803	0.331	-5.452	**0.0001**
LanguageY:ToneB	-0.576	0.333	-1.728	0.0888

***ToneSpaceOffglide3:* Mandarin vs. Thai and Yoruba**				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageT	1.179	2.711	0.435	0.534
LanguageY	-1.564	2.711	-0.577	0.4046
ToneB	-6.387	0.253	-25.259	**0.0001**
LanguageT:ToneB	-1.883	0.346	-5.435	**0.0001**
LanguageY:ToneB	-0.657	0.349	-1.883	0.0634

***ToneSpaceOffglide4:* Thai vs. Yoruba**				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageY	-2.742	2.199	-1.25	0.0878
ToneB	-8.255	0.22	-37.55	**0.0001**
LanguageY:ToneB	1.2312	0.313	3.93	**0.0002**

Table 4.6. Summary of the results of the *ToneSpaceOffglide* lmers

The above values are significant at $p \leq 0.0125$ (after Bonferroni correction). The results of *ToneSpacesOffglide* are as follows:

1. There is no main effect of language. The Grand Mean F0 did not differ as a function of language.
2. Overall, there is a significant difference between the top and bottom tones. In each of the models, the bottom tone was 6.3-8.3 ST lower on average than the top tone. This indicates that the top and bottom tones are well-differentiated overall.
3. The interaction of tone and language is significant only in comparisons involving Thai.

The results of *ToneSpaceOffglide* corroborate observations of the data in Figure 4.5. There was no main effect of language, and the tone x language interaction results indicate the following hierarchy with regard to tone-space size at tonal offglide: Igbo = Yoruba = Mandarin = Cantonese < Thai.

The *ToneSpaceOffglide* results indicate, per hypothesis H1, that tone-space-size is fixed across languages at tonal offglide, with one exception (Thai). It is possible that Thai requires an expanded tone space at tonal offglide to differentiate its R and H tones at offglide, as both end at the top of the tonal space, and both have F0 trajectories that flatten out from timepoint 8 to the offglide. Both the results of *ToneSpaceOffglide* and those of *ToneSpace* may support reports in the literature that pitch excursion in Thai is of greater importance than offglide for tone perception (Pike 1948; Gandour 1983; and others). In other words, it may not matter what the tonal F0 at offglide is for Thai tone perception, so long as the F0 excursion across the tone trajectory is distinctive.

4.3.4. Summary of *ToneSpace* analyses

The flowchart in Figure 4.6 illustrates the results of the *ToneSpace* analyses.

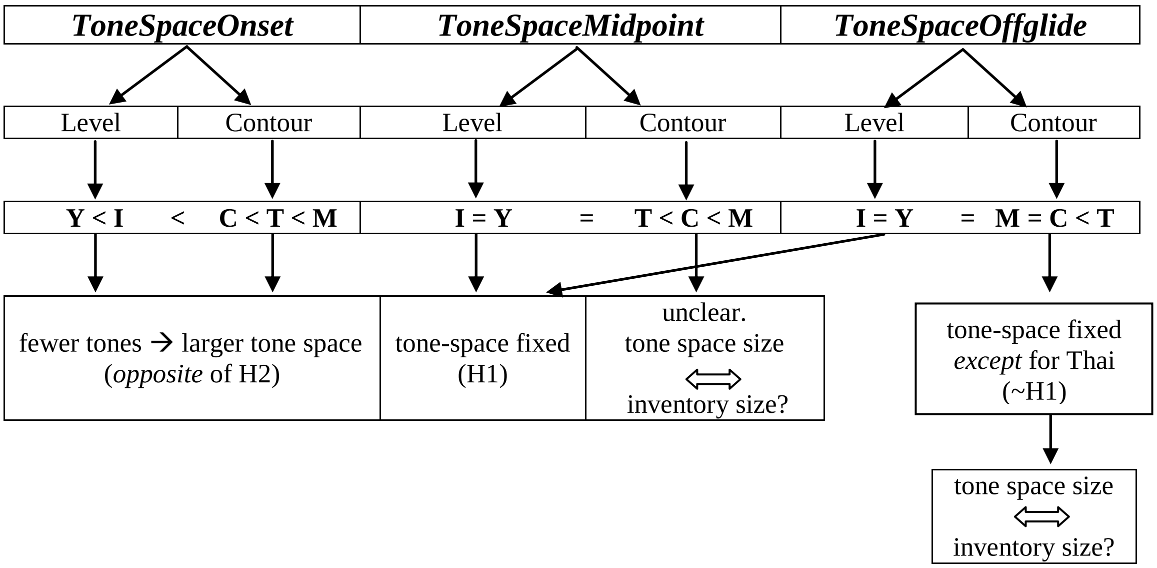


Figure 4.6. Flowchart summarizing the *ToneSpace* analyses

One of the key trends illustrated in the flowchart is that tone-space size appears to be fixed across level-tone languages at midpoint and offglide. Tone-space size was fixed at offglide across the contour-tone languages as well, with the exception of Thai. Additionally, within each language type (level vs. contour), languages with smaller tone inventories had larger tone spaces at tonal onset. Finally, the results of the *ToneSpace* analyses at midpoint and offglide lead to the question as to whether or not tone-space size is genuinely correlated with inventory size.

4.4. Examination of cross-language tone dispersion

In this section I present and test, via a series of *ToneDisp* models, TAD predictions of cross-language tone dispersion. In each *ToneDisp* lmer, I compare the F0 difference between a baseline tone (the H tone) and a second tone that is especially phonetically-similar across languages. Across languages, the second tone is considered comparatively further dispersed from the baseline if the F0 difference between the H and second tone is larger. The corrected α-level here and in all other analyses was calculated as 0.05/[number of lmers]. As the number of lmers is 2 for each analysis in this section, significance is assessed at the 0.05% level.

Referring back to Figures 2.1-2.5, it is apparent that the H and M level tones in Cantonese and Yoruba are strikingly similar with respect to their F0 trajectories. The H and R tone in Mandarin and the H and MR tones in Cantonese are notably similar in this same respect. In the ToneDispH-M models that follow, I compare the H-M tone F0 difference between Cantonese and Yoruba. In the *ToneDispH-R* models, I compare the H-R tone F0 difference in Mandarin with the H-MR difference in Cantonese.

Recall that I intended to test the hypotheses and predictions outlined on pages 64-65. However, the predictions for cross-language tone dispersion under H1 were predicated upon finding that Cantonese = Thai = Mandarin = Yoruba = Igbo in overall tone-space size. Similarly, the predictions for cross-language tone dispersion under H2 were predicated upon finding that Cantonese > Thai > Mandarin > Yoruba > Igbo in overall tone-space size. The *ToneSpace* analyses yielded neither of these outcomes. So, the aforementioned predictions for cross-language tone dispersion must be modified in favor of reformulated predictions that follow directly from the results of the *ToneSpace* models. The general principle behind the new predictions remains the same, namely, the degree of tonal dispersion displayed by a language is correlated with both its tone-space size and the size of its tonal inventory. However, the reformulated predictions neither assume that the size of the tone space is positively correlated with the size of the tone inventory, nor that tone-space size is fixed across languages. The reformulated predictions instead take into account the Cantonese, Mandarin, and Yoruba tone space sizes that were determined by the *ToneSpace* analyses. Importantly, they assume (in keeping with H2, as schematized in Figure 4.2) that tones located within the tone space will be evenly dispersed within the space.

Figure 4.7 is an illustration of the tone-spaces and predicted degree of M tone dispersion, relative to the H tone baseline, of Yoruba vs. Cantonese at tonal onset, midpoint, and offglide. Figure 4.8 is an illustration of the tone-spaces and predicted degree of R/MR tone dispersion, relative to the H tone baseline, of Cantonese vs. Mandarin at tonal onset, midpoint, and offglide. Tone-space size is represented as the distance between the top and bottom tone in each language, per the *ToneSpace* analyses, and are sized to scale based on the *ToneSpace* analysis results present in section 4.3 (rounded to the nearest whole ST). The Cantonese tone space spans 6 ST at onset, 8 ST at midpoint, and 7 ST at offglide; the Yoruba tone space spans 3 ST at onset, 6 ST at midpoint, and 7 ST at offglide; and the Mandarin tone space spans 8 ST at onset, 9 ST at midpoint, and 7 ST at offglide. Calculations and explanations of these tone spaces are provided in *Appendix D: Calculations of Cantonese, Mandarin, and Yoruba Tone-space Sizes at Onset, Midpoint, and Offglide for Section 4.4*.

Following the TAD, the tones of each language are considered to be repellers in a dynamical system, so each is located maximally far from the others; inter-tonal distance is the Euclidean distance in F0 (ST). The top tone is the H tone, as shown. The curly brackets in Figure 4.7 indicate the predicted distance between the H and M tones in Yoruba vs. Cantonese, and the curly brackets in Figure 4.8 indicate the predicted distance between the H and R tones in Mandarin vs. between the H and MR tones in Cantonese. The locations of the M, R, and MR tones in the idealized spaces in Figures 4.7 and 4.8 reflect their locations in Figures 2.1-2.5.

Because its F0 drops over the course of its trajectory, the M tone is the second-highest tone in the Cantonese tone space at onset; it is third-highest at midpoint, and fourth-highest at offglide. In contrast, the Yoruba M tone is in the middle of its tone space at all three timepoints. The Cantonese MR tone and the Mandarin R tone are both the second-highest tones in their respective tone spaces at all three timepoints. The estimated degree of dispersion of the M, R, or MR tone relative to the H tone is the overall tone space size divided by the total number of inter-tone intervals in the space, multiplied by the number of inter-tone intervals between the H tone and the M, R, or MR tone. For instance, the Cantonese MR tone at offglide is estimated to be about 4.2 ST dispersed from the H tone by the following equation: ((7 [tone-space size in ST]) / (5 [number of total inter-tonal intervals within the tone space])) x (3 [number of inter-tonal intervals between the H tone and MR tone]) = 4.2 ST.

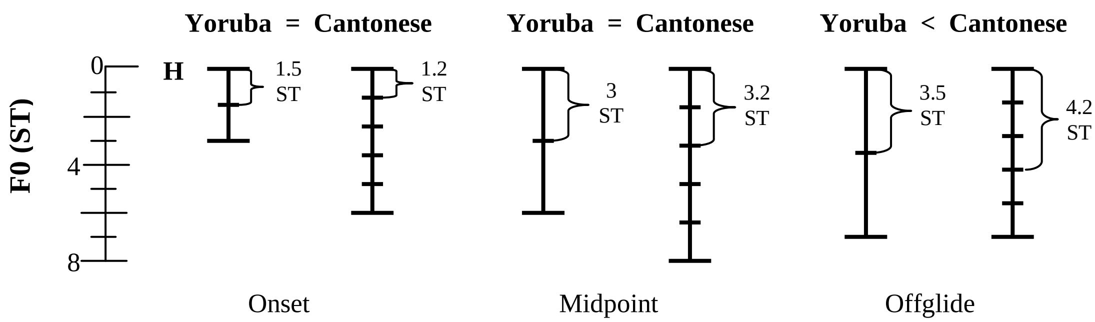


Figure 4.7. An idealized illustration of the Yoruba and Cantonese tone-space areas and predicted relative degrees of M-tone dispersion within the tone spaces

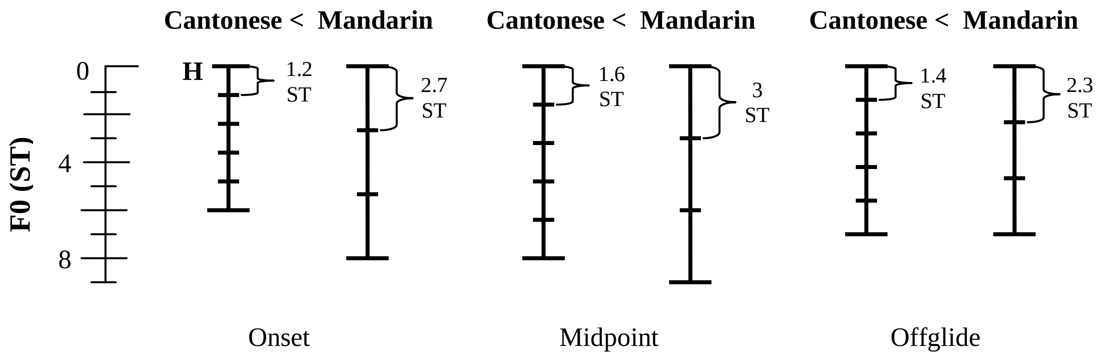


Figure 4.8. An idealized illustration of the Cantonese and Mandarin tone-space areas and predicted relative degrees of R/MR-tone dispersion within the tone spaces

As indicated in Figures 4.7-4.8, I predict the following tonal dispersion hierarchies:

ToneDispH-M:

- *ToneDispH-MOnset* and *Midpoint*: **Yoruba = Cantonese**
- *ToneDispH-MOffglide*: **Yoruba < Cantonese**

ToneDispH-R Onset, *Midpoint*, and *Offglide*: **Cantonese < Mandarin**

Note that Figure 4.7 indicates that the H-M distance in Yoruba is greater than that of Cantonese at onset, and that the H-M distance in Cantonese is greater than that of Yoruba at midpoint. However, I predict Yoruba and Cantonese will display equivalent degrees of M-tone dispersion at those two timepoints. The differences between the Yoruba and Cantonese values at those timepoints are negligible (0.2-0.3 ST), and those values are approximated, not precise.

4.4.1. *ToneDispH-M*

4.4.1.1. *ToneDispH-MOnset*

ToneDispH-MOnset compares the degree of dispersion between the H baseline tone and the M tone in the Cantonese and Yoruba tone spaces at tonal onset. Figure 4.9 shows the H-M tone Mean F0 at timepoint k1 for the languages. Standard error bars surround each data point.

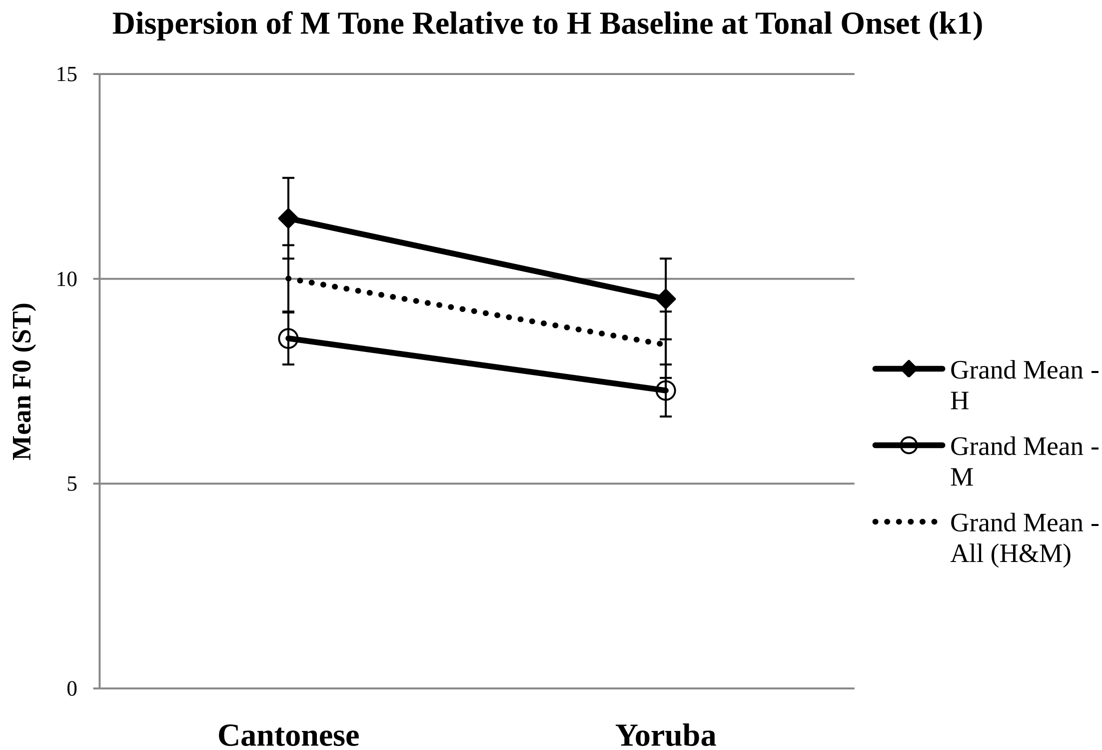


Figure 4.9. H - M F0 differences (in ST) in Cantonese and Yoruba at tonal onset

Item	Cantonese	Yoruba
Grand Mean - H	11.477	9.506
Grand Mean - M	8.542	7.272
Grand Mean – All (H&M)	10.009	8.389
Grand Mean -- H-M	2.935	2.234

Table 4.7. M-tone dispersion F0 (ST) values in Cantonese and Yoruba at tonal onset

Note that the Cantonese Grand Mean – All (H&M) value is approximately 1.6 ST higher than that of Yoruba. Additionally, the difference between the languages' H – M Grand Mean F0 values is only about 0.7 ST. Table 4.8, below, summarizes the fixed-effects results of the *ToneDispH-MOnset* lmer.

***ToneDispH-MOnset:* Cantonese vs. Yoruba**				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageY	-2.086	2.576	-0.810	**0.0084**
ToneM	-2.925	0.089	-32.860	**0.0001**
LanguageY:ToneM	0.661	0.100	6.620	**0.0001**

Table 4.8. Summary of the results of the *ToneDispH-MOnset* lmer

The results of *ToneDispH-MOnset* are as follows:

1. <u>There is a main effect of language.</u> Grand Mean F0 differed as a function of language.
2. <u>Overall, there is a significant difference between the H and M tones.</u> The M tone was 2.9 ST lower on average than the H tone. This indicates that the H and M tones are well-differentiated overall.
3. <u>The interaction of tone and language was significant.</u>

The results of *ToneDispH-MOnset* are slightly surprising. The tone x language interaction results indicate that Yoruba (3 tones) < Cantonese (6 tones) with regard to the difference between the H and M tones' mean F0 at tonal onset, despite there being less than 1 ST difference between the two languages' H-M Grand Mean F0 values.

Recall that I predicted to find that Yoruba (3 tones) = Cantonese (6 tones) with regard to dispersion of the M tone from the H tone in the tone space at tonal onset. The results of *ToneDispH-MOnset* do not support this expectation.

4.4.1.2. ***ToneDispH-MMidpoint***

ToneDispH-MMidpoint compares the degree of dispersion between the H baseline tone and the M tone in the Cantonese and Yoruba tone spaces at tonal midpoint. Figure 4.10 shows the H-M tone Mean F0 at timepoint k5 for the languages. Standard error bars surround each data point.

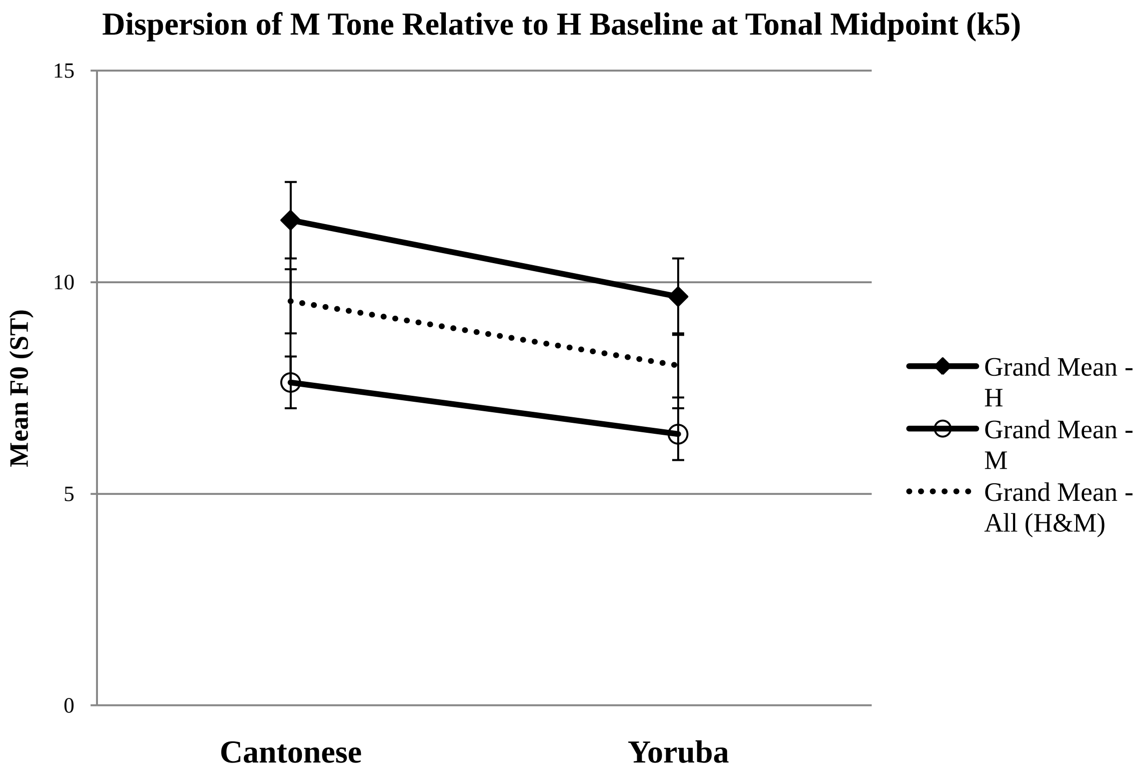


Figure 4.10. H - M F0 differences (in ST) in Cantonese and Yoruba at tonal midpoint

Item	Cantonese	Yoruba
Grand Mean - H	11.466	9.659
Grand Mean - M	7.633	6.409
Grand Mean – All (H&M)	9.549	8.034
Grand Mean -- H-M	3.833	3.250

Table 4.9. M-tone dispersion F0 (ST) values in Cantonese and Yoruba at tonal midpoint

The two languages' Grand Mean – All (H&M) values differ by about 1.5 ST. Additionally, the difference between the languages' H – M Grand Mean F0 values is about 0.6 ST. Table 4.10, below, summarizes the fixed-effects results of the *ToneDispH-MMidpoint* lmer.

ToneDispH-MMidpoint: **Cantonese vs. Yoruba**				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageY	-1.905	2.534	-0.750	**0.028**
ToneM	-3.828	0.110	-34.69	**0.0001**
LanguageY:ToneM	0.5453	0.120	4.550	**0.0001**

Table 4.10. Summary of the results of the *ToneDispH-MMidpoint* lmer

The results of *ToneDispH-MMidpoint* are as follows:

1. <u>There is a main effect of language</u>. Grand Mean F0 differed as a function of language.
2. <u>Overall, there is a significant difference between the H and M tones.</u> The M tone was about 2.3 ST lower on average than the H tone. This indicates that the H and M tones are well-differentiated overall.
3. <u>The interaction of tone and language was significant.</u>

The results of *ToneDispH-MMidpoint* are slightly surprising as well. The tone x language interaction results indicate that Yoruba (3 tones) < Cantonese (6 tones) with regard to the difference between the H and M tones' mean F0 at tonal midpoint, despite there being less than a 1 ST difference between the two languages' H-M Grand Mean F0 values.

Recall that I predicted that Yoruba = Cantonese with regard to degree of dispersion of the M tone from the H tone in the tone space at tonal midpoint. The results of *ToneDispH-MMidpoint* do not support this expectation.

4.4.1.3. *ToneDispH-MOffglide*

ToneDispH-MOffglide compares the degree of dispersion between the H baseline tone and the M tone in the Cantonese and Yoruba tone spaces at tonal offglide. Figure 4.11, below, shows the H-M tone Mean F0 at timepoint k9 for the languages. Standard error bars surround each data point.

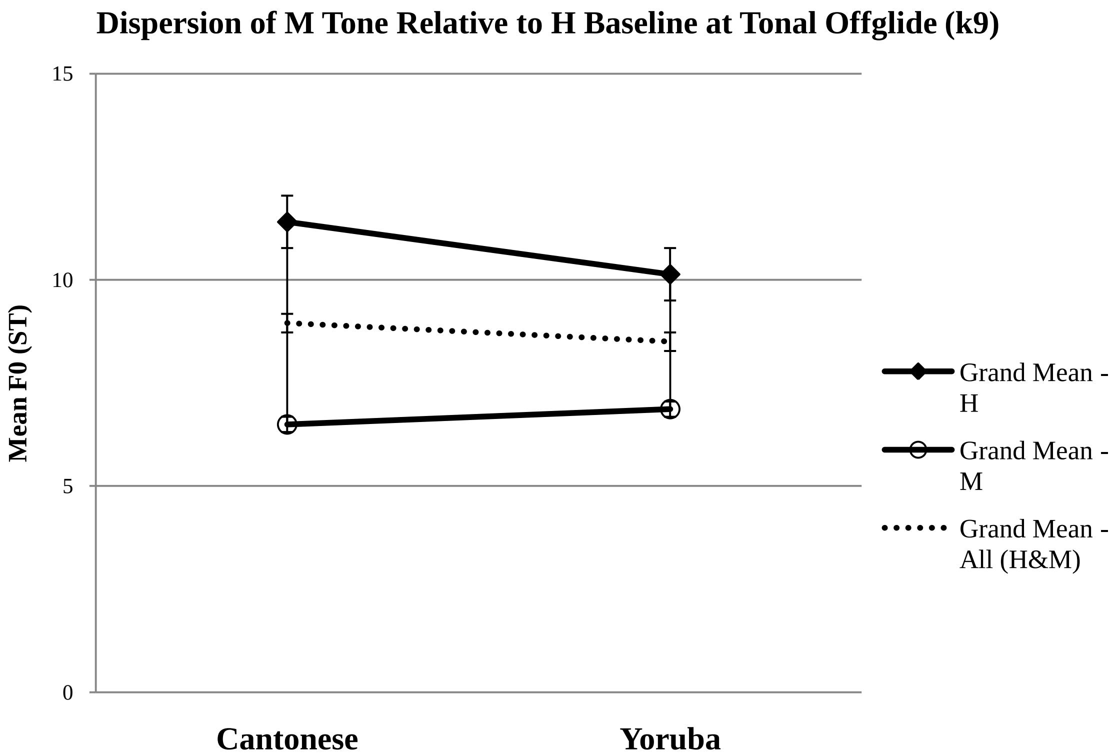


Figure 4.11. H - M F0 differences (in ST) in Cantonese and Yoruba at tonal offglide

Item	Cantonese	Yoruba
Grand Mean - H	11.405	10.134
Grand Mean - M	6.493	6.863
Grand Mean – All (H&M)	8.949	8.498
Grand Mean -- H-M	4.912	3.271

Table 4.11. M-tone dispersion F0 (ST) values in Cantonese and Yoruba at tonal offglide

Note that the Cantonese Grand Mean – All (H&M) value is only about 0.45 ST higher than that of Yoruba. Additionally, the Yoruba Grand Mean H-M F0 difference is about 1.6 ST smaller than that of Cantonese. Table 4.12, below, summarizes the fixed-effects results of the *ToneDispH-MOffglide* lmer.

ToneDispH-MOffglide: **Cantonese vs. Yoruba**				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageY	-1.336	2.522	-0.53	0.4358
ToneM	-4.895	0.248	-19.77	**0.0001**
LanguageY:ToneM	1.5738	0.305	5.161	**0.0001**

Table 4.12. Summary of the results of the *ToneDispH-MOffglide* lmer

The results of *ToneDispH-MOffglide* are as follows:

1. <u>There is no main effect of language.</u> Grand Mean F0 did not differ as a function of language.
2. <u>Overall, there is a significant difference between the H and M tones.</u> The M tone was nearly 5 ST lower on average than the H tone. This indicates that the H and M tones are well-differentiated overall.
3. <u>The interaction of tone and language was significant.</u>

The results of *ToneDispH-MOffglide* corroborate observations of the data in Figure 4.11. There was no main effect of language and, importantly, the tone x language interaction results indicate that Yoruba (3 tones) < Cantonese (6 tones) with regard to the difference between the H and M tones' mean F0 at tonal offglide.

Recall that I predicted that Yoruba < Cantonese with regard to degree of dispersion of the M tone from the H tone in the tone space at tonal offglide. The results of *ToneDispH-MOffglide* support this expectation.

4.4.1.4. Summary of *ToneDispH-M* results

In all the models, the H and M tones were significantly separated overall, with the M tone being lower than that of the H tone. Additionally, Yoruba (3 tones) < Cantonese (6 tones) with regard to the dispersion of the M from the baseline H tone at all three timepoints. Only one prediction was supported by the analyses above: *ToneDispH-MOffglide* showed that Yoruba (3 tones) < Cantonese (6 tones) with regard to M dispersion from baseline at offglide. It appears as though the M tone of the language with the larger tone inventory (Cantonese) is more dispersed from the H baseline than that of the language with the smaller inventory (Yoruba) across the tonal trajectory. Not only does Cantonese have the larger tone space at onset and midpoint, it also displays greater tone dispersion relative to Yoruba. This is inconsistent with the TAD: having both an expanded overall tone space *and* greater tone dispersion would be considered inefficient and theoretically unnecessary.

4.4.2. *ToneDispH-R*

4.4.2.1. *ToneDispH-ROnset*

ToneDispH-ROnset compares the degree of dispersion between the H baseline tone and the Mandarin R tone or Cantonese MR tone at tonal onset. Figure 4.12 shows the H-R/MR tone mean F0 at timepoint k1 for the languages. Standard error bars surround each data point.

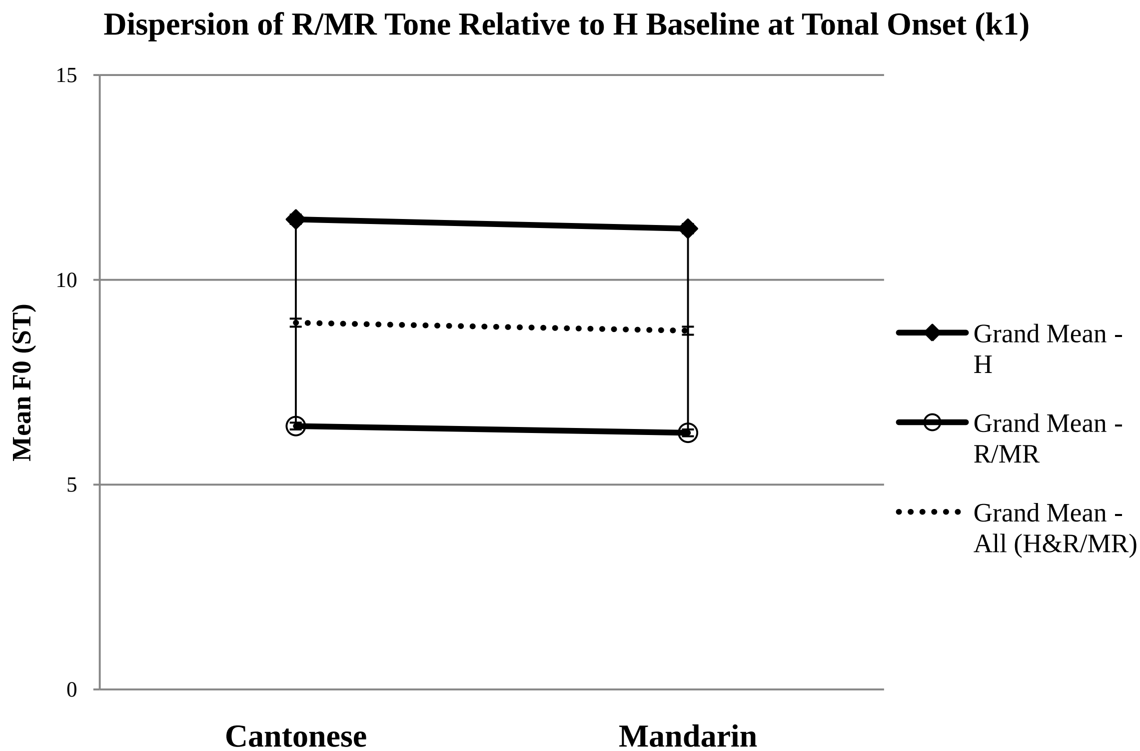


Figure 4.12. H – R/MR F0 differences (in ST) in Cantonese and Mandarin at tonal onset

Item	Mandarin	Cantonese
Grand Mean - H	11.249	11.477
Grand Mean – R/MR	6.263	6.429
Grand Mean – All (H&R/MR)	8.756	8.953
Grand Mean -- H-R/MR	2.493	2.524

Table 4.13. F0 of H – R/MR tones in Cantonese and Mandarin at tonal onset

Both the Grand Mean – All (H&R/MR), and the H-R/MR F0, values are nearly the same across the languages. Table 4.14, below, summarizes the fixed-effects results of *ToneDispH-ROnset*.

***ToneDispH-R Onset:* Cantonese vs. Mandarin**				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageM	-0.254	2.943	-0.090	0.7418
ToneMR	-5.033	0.093	-53.840	**0.0001**
LanguageM:ToneR	0.049	0.108	0.460	0.673

Table 4.14. Summary of the results of the *ToneDispH-ROnset* lmer

The results of *ToneDispH-ROnset* are as follows:

1. <u>There is no main effect of language.</u> Grand Mean F0 did not differ as a function of language.
2. <u>Overall, there is a significant difference between the H and R tones.</u> The R/MR tone was about 5 ST lower on average than the H tone. This indicates that the H and R/MR tones are well-differentiated overall.
3. <u>The interaction of tone and language was not significant.</u>

ToneDispH-ROnset supports observations of the data in Figure 4.12. There is no main effect of language, and the tone x language interaction suggests that Mandarin (4 tones) = Cantonese (6 tones) with regard to the dispersion of the R or MR tone from the baseline H tone at tonal onset.

Recall that I predicted that Cantonese < Mandarin with regard to dispersion of the R/MR tone from the H tone in the tone space at tonal onset. The results of *ToneDispH-ROnset* do not support this expectation.

4.4.2.2. *ToneDispH-RMidpoint*

ToneDispH-RMidpoint compares the degree of dispersion between the H baseline tone and the Mandarin R tone or Cantonese MR tone at tonal midpoint. Figure 4.13 shows the H-R/MR tone mean F0 at timepoint k5 for the languages. Standard error bars surround each data point.

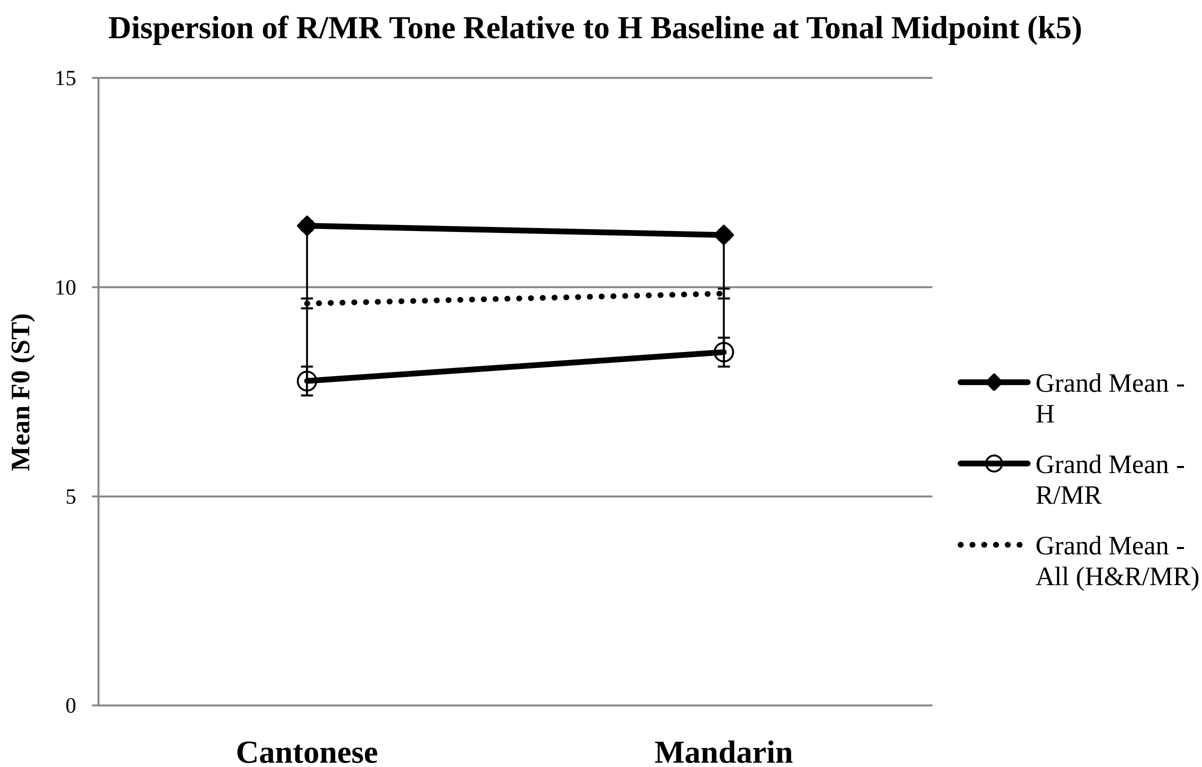


Figure 4.13. H - R/MR F0 differences (in ST) in Cantonese and Mandarin at tonal midpoint

Item	Cantonese	Mandarin
Grand Mean - H	11.466	11.247
Grand Mean – R/MR	7.756	8.446
Grand Mean – All (H&R/MR)	9.611	9.847
Grand Mean -- H-R/MR	3.710	2.801

Table 4.15. F0 of H – R/MR tones in Cantonese and Mandarin at tonal midpoint

Observe that the Cantonese Grand Mean – All (H&R/MR) value is only about 0.2 ST smaller than that of Mandarin. Also, the languages' Grand Mean H-R/MR F0 values differ by less than 1 ST. Table 4.16, below, summarizes the fixed-effects results of *ToneDispH-RMidpoint*.

ToneDispH-RMidpoint: **Cantonese vs. Mandarin**				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageM	-0.250	2.780	-0.090	0.7148
ToneMR	-3.670	0.097	-37.890	**0.0001**
LanguageM:ToneR	0.883	0.093	9.490	**0.0001**

Table 4.16. Summary of the results of the *ToneDispH-RMidpoint* lmer

The results of *ToneDispH-RMidpoint* are as follows:

1. <u>There is no main effect of language.</u> Grand Mean F0 did not differ as a function of language.
2. <u>Overall, there is a significant difference between the H and R tones.</u> The R/MR tone was about 3.7 ST lower on average than the H tone. This indicates that the H and R/MR tones are well-differentiated overall.
3. <u>The interaction of tone and language was significant.</u>

ToneDispH-RMidpoint corroborates the data in Figure 4.13. There was no main effect of language, and the tone x language interaction indicates that Mandarin (4 tones) < Cantonese (6 tones) with regard to the dispersion of the R/MR from the baseline H tone at tonal midpoint.

I had predicted that Cantonese < Mandarin with regard to degree of dispersion of the R/MR tone from the H tone in the tone space at tonal midpoint. The results of *ToneDispH-RMidpoint* do not support this prediction.

4.4.2.3. *ToneDispH-ROffglide*

ToneDispH-ROnset compares the degree of dispersion between the H baseline tone and the Mandarin R tone or Cantonese MR tone at tonal onset. Figure 4.14, below, shows the H-R/MR tone mean F0 at timepoint k9 for the languages. Standard error bars surround each data point. Note that the scale on the y-axis is much smaller than that of previous figures (namely, 11-12 ST, in 0.5-ST increments), to make Grand Mean differences visible.

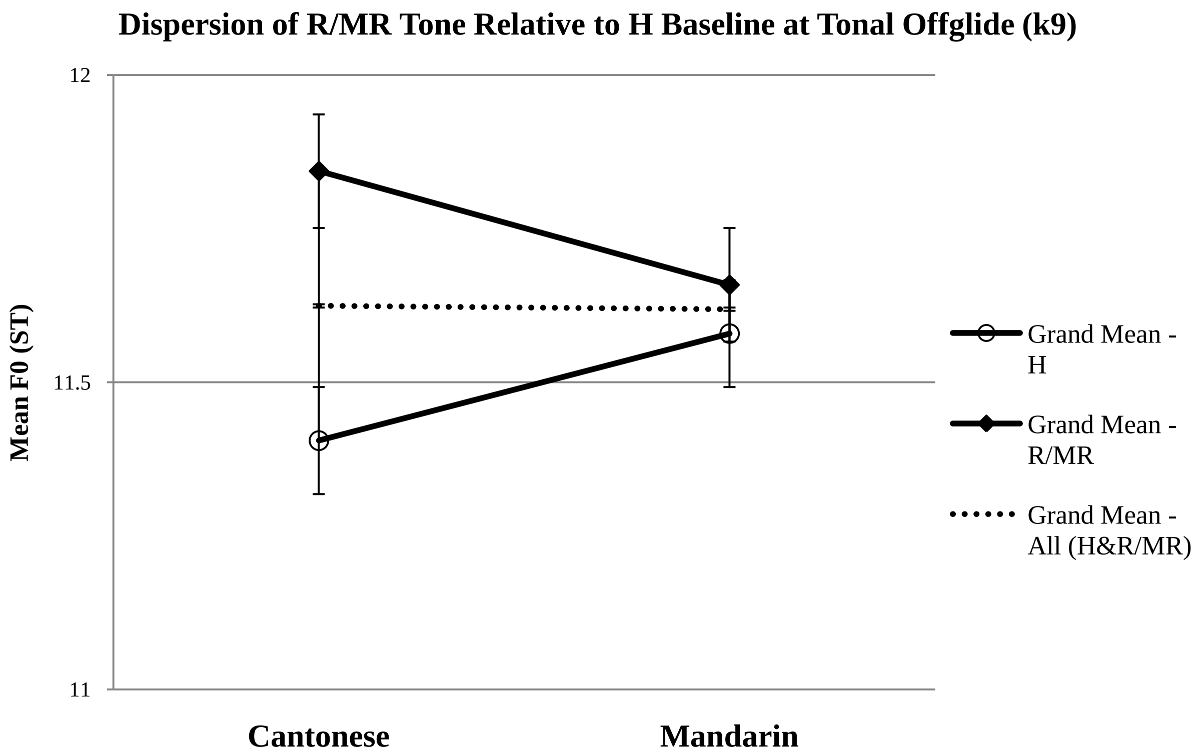


Figure 4.14. H - R/MR F0 differences (in ST) in Cantonese and Mandarin at tonal offglide

Item	Cantonese	Mandarin
Grand Mean - H	11.405	11.579
Grand Mean – R/MR	11.843	11.658
Grand Mean – All (H&R/MR)	11.624	11.619
Grand Mean -- H-R/MR	-0.438	-0.079

Table 4.17. F0 of H – R/MR tones in Cantonese and Mandarin at tonal offglide

The Grand Mean -- All (H&R/MR) value is almost identical across the languages, and the F0 difference between the Mandarin H and R tones is only 0.4 ST smaller than the F0 difference between the Cantonese H and MR tones. Table 4.18, below, summarizes the fixed-effects results of the *ToneDispH-ROffglide* lmer.

***ToneDispH-ROffglide:* Cantonese vs. Mandarin**				
	Est	**St.E**	**t-val**	**pMCMC**
LanguageM	0.077	3.010	0.026	0.9568
ToneMR	0.449	0.294	1.530	0.1264
LanguageM:ToneR	-0.339	0.369	-0.918	0.3622

Table 4.18. Summary of the results of the *ToneDispH-ROffglide* lmer

The results of *ToneDispH-ROffglide* are as follows:

1. <u>There is no main effect of language.</u> Grand Mean F0 did not differ as a function of language.
2. <u>Overall, the difference between the H and R/MR tones is not significant.</u> The H and R/MR tones are not significantly well-differentiated in the languages at offglide.
3. <u>The interaction of tone and language was not significant.</u>

ToneDispH-ROffglide corroborates observations of the data in Figure 4.14. There was no main effect of language and, as in *ToneDispH-ROnset*, the tone x language interaction indicates that Mandarin (4 tones) = Cantonese (6 tones) regarding the dispersion of the R from the baseline H tone at tonal offglide.

I had predicted that Cantonese (6 tones) < Mandarin (4 tones) with regard to degree of dispersion of the R/MR tone from the H tone in the tone space at tonal offglide. The results of *ToneDispH-RMidpoint* did not support this prediction.

4.4.2.4. Summary of *ToneDispH-R* results

The crucial result of these models was in regard to the degree of dispersion of the Mandarin R or Cantonese MR tone relative to the baseline H tone (in Grand Mean F0, in ST). *ToneDispH-ROnset* and *Offglide* indicated that Mandarin (4 tones) = Cantonese (6 tones) at tonal onset and offglide; and *ToneDispH-RMidpoint* indicated that Mandarin (4 tones) < Cantonese (6 tones) at midpoint. None of these results supported the predictions made earlier.

4.4.3. Summary of *ToneDisp* results

The flowchart in Figure 4.15 illustrates the results of the *ToneDisp* analyses.

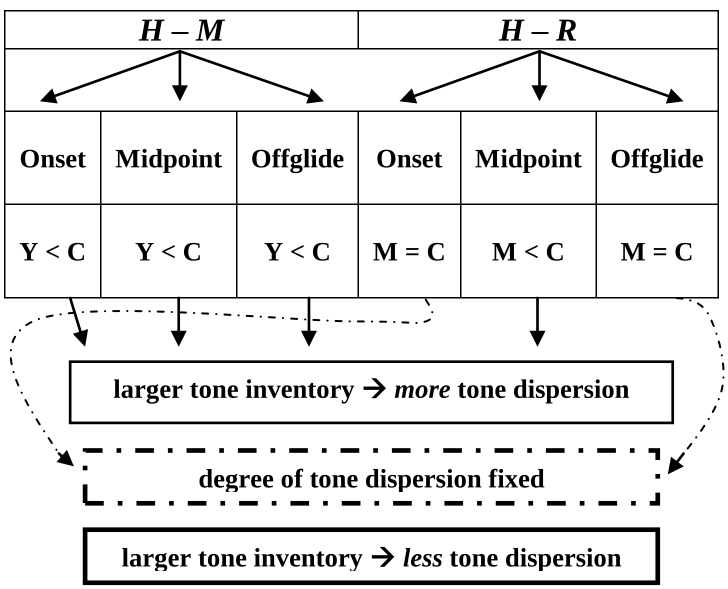


Figure 4.15. Flowchart summarizing the *ToneDisp* analyses

One of the key trends illustrated in the above flowchart is that only one of the predictions made in this section were supported by the data: the results of *ToneDispH-MOffglide* indicated, as predicted, that Yoruba < Cantonese with regard to degree of M-tone dispersion relative to the H-tone baseline. In fact, Yoruba < Cantonese in M-tone dispersion at onset and midpoint as well. Cantonese has both a relatively expanded tone space *and* greater tonal dispersion at onset and midpoint (recall that at offglide, Yoruba = Cantonese in tone-space size). This is inconsistent with the TAD: having both an expanded overall tone space *and* greater tone dispersion is inefficient and theoretically unnecessary.

Because Cantonese (6 tones) < Mandarin (4 tones) with regard to tone-space size at onset and midpoint, I predicted that Cantonese < Mandarin in H – R/MR tone dispersion at onset and midpoint. However, relative degree of tone dispersion at those timepoints was such that Mandarin = Cantonese at onset (meaning that the tones of Cantonese are comparatively overly crowded at onset) and Mandarin < Cantonese at midpoint (meaning that the tones of Cantonese are theoretically overly dispersed at midpoint). At offglide, Mandarin = Cantonese in tone-space size, and I predicted that the language with the larger tone inventory (Cantonese) < the language with the smaller tone inventory (Mandarin) in degree of H – R/MR tone dispersion at offglide. However, Mandarin = Cantonese in tonal dispersion at offglide. This result is also inconsistent with the TAD: the tones of the language with the larger inventory (Cantonese) would be expected to be more crowded than those of the language with the smaller tone inventory (Mandarin) if their overall tone-spaces are equivalent in size.

The results of the *ToneSpace* and *ToneDisp* analyses showed that the Theory of Adaptive Dispersion cannot adequately account for the cross-language tone-space and tone-dispersion data presented in this study. In chapter five, I briefly recap the overview and results of this study, offer conjectures as to what might more accurately account for the current data, conduct some additional analyses, and provide suggestions for future work.

CHAPTER FIVE: CONCLUSIONS AND DISCUSSION

5.1. General overview of the study

The research presented in this dissertation was motivated by a general interest in the possible effect of tone inventory size and composition on acoustic tone-space size and tonal dispersion. This interest arises in large part from the observation that, while about 42% of the world's languages are tonal (Maddieson, 2008), and more than 50% of the world's population speak a tone language (Fromkin, 1978), tone languages are under-studied compared to segmental contrast systems. This study is also motivated by an interest in discovering whether or not well-studied models of segmental (vowel) system organization (in particular, the TAD) accurately predict tone-system configurations. The current study was therefore designed to test, for five languages with very different tone-system configurations, specific hypotheses and predictions of the TAD. The tone systems of Cantonese (3 contour tones, 3 level tones), Mandarin (3 contour tones, 1 level tone), Thai (2 contour tones, 3 level tones), Yoruba (3 level tones), and Igbo (2 level tones only) were examined in order to determine whether and how (a) the overall size of the acoustic tone space differs across languages as a function of tone-inventory size; and (b) dispersion of tone categories within the tone space differs across languages as a function of tone-space and tone-inventory size.

In this chapter, I first briefly recap the results of the *ToneSpace* and *ToneDisp* analyses. I then conduct some alternative analyses, offer conjectures as to what might also more accurately account for the current data, and ultimately describe experiments needed to test those accounts.

5.2. Brief recap of results

The first goal of this study was to examine and compare the languages' overall tone-space areas. Specifically, I attempted to determine if and how the languages' tone space sizes differ from one another as a function of the composition (number and type) of tones in their inventories. To this end, I tested two competing hypotheses and their accompanying predictions. H1 states that the size of the acoustic tone space is independent of the size of the tone inventory. With regard to the languages under investigation, H1 led to the prediction that Cantonese (6 tones) = Thai (5 tones) = Mandarin (4 tones) = Yoruba (3 tones) = Igbo (2 tones) in overall tone-space size. H2 states that the size of the acoustic tone space is positively correlated with tone inventory size. H2 led to the prediction that Cantonese (6 tones) > Thai (5 tones) > Mandarin (4 tones) > Yoruba (3 tones) > Igbo (2 tones) in overall tone-space size. I defined tone-space size as the Grand Mean F0 difference between each language's highest (top) and lowest (bottom) tones, and found the following hierarchies of tone-space sizes:

- Yoruba < Igbo < Cantonese < Thai < Mandarin at tonal onset
- Igbo = Yoruba = Thai < Cantonese < Mandarin at midpoint
- Igbo = Yoruba = Mandarin = Cantonese < Thai at offglide

The second goal of this study was to investigate whether and how the dispersion of phonetically-similar tone categories within the tone space differed across languages as a function of the size of their tone spaces and tone inventories. I compared the F0 difference between a baseline tone (the H tone, shared across languages) and (a) the M tone in Cantonese vs. Yoruba and (b) the Cantonese MR tone vs. the Mandarin R tone. The M or R/MR tone was considered comparatively further dispersed from the baseline if the F0 difference between it and the H tone was larger. Following the TAD, I assumed that tone categories act as repellers in a dynamical system, and would find equilibrium when located far from other tone categories. Crucially, I tested hypotheses and predictions that followed directly from the results of the *ToneSpace* analyses. That is, they take into account the Cantonese, Mandarin, and Yoruba tone space sizes that were determined by the *ToneSpace* analyses as well as where each tone was located in the tone space at onset, midpoint, and offglide (from Figures 2.1, 2.3, and 2.4). Table 5.1 displays the *ToneDisp* predictions and results. Analyses of tonal dispersion at onset are listed first, followed by analyses of dispersion at midpoint and offglide. Note that the predictions matched the results of only one analysis, *ToneDispH-ROffglide* (highlighted in boldface font below).

Analysis	**Prediction**	**Results**
*ToneDispH-M**Onset***	Yoruba = Cantonese	Yoruba < Cantonese
*ToneDispH-R**Onset***	Cantonese < Mandarin	Mandarin = Cantonese
*ToneDispH-M**Midpoint***	Yoruba = Cantonese	Yoruba < Cantonese
*ToneDispH-R**Midpoint***	Cantonese < Mandarin	Mandarin < Cantonese
*ToneDispH-M**Offglide***	**Yoruba < Cantonese**	**Yoruba < Cantonese**
*ToneDispH-R**Offglide***	Cantonese < Mandarin	Mandarin = Cantonese

Table 5.1. Results of the *ToneDisp* lmers

5.3. Discussion

The results of the *ToneSpace* and *ToneDisp* analyses shows that the Theory of Adaptive Dispersion does not adequately account for the cross-language tone-space and tone-dispersion data presented in this study. However, this is not entirely surprising. Recall that multiple studies on vowel systems found that, counter to predictions of the TAD, larger vowel inventories had larger vowel spaces (e.g., Gendrot and Adda-Decker, 2007). Likewise, various studies on vowel dispersion found that vowels are not always dispersed evenly across their vowel spaces (e.g., Disner, 1983).

Gendrot and Adda-Decker (2007), in a comparison of the vowel systems of English, French, German, Italian, Mandarin Chinese, Portuguese, and Spanish, found that languages with larger vowel inventories did not have expanded vowel spaces. The authors suggested that a negative result such as theirs could be interpreted to mean that other acoustic and/or articulatory dimensions are used to distinguish otherwise-similar vowels. For instance, the nasality in, e.g., French vowels, may be used for this purpose; diphthongization, voice quality, and voicing may be other such mechanisms that are employed. The same is likely true in tone languages. Though F0 is considered to be the primary acoustic correlate for the languages examined in this study, many may – or are known to – use other acoustic correlates to help distinguish their tones. Though not measured, the Igbo recordings in this study indicated that amplitude may be a secondary cue to tone identity: the amplitude of the L tone was informally observed to be consistently lower across talkers. Similarly (and again not measured), the Yoruba recordings of the current study indicated that voice quality may be a cue to L tone identity: male and female speakers alike consistently produced the L tone with a breathy voice quality. Other phonetic correlates of Mandarin tones include syllable amplitude (Gårding et al., 1986); the shape of the amplitude envelope (Fu et al., 1998); voice quality (Gårding et al., 1986), e.g., creak (glottalization) along the FR tone trajectory; and temporal properties such as duration (e.g., the F tone is typically shortest and the FR tone longest in duration) and Turning Point (Lin, 1965; Chuang et al., 1972; Jongman and Moore, 2000; Fu and Zheng, 2000; Blicher, et al., 1990, and others). In Thai, the phonetic shapes of the individual tones – even in citation form – do not match their labels well (Abramson, 1962; Gandour et al., 1991; Zsiga and Nitisaroj, 2007); as such, Turning Point may be crucial to Thai tone identity as well. In addition, phonation type may be important for Thai tone identity, as F and H tones are produced with creak (Wayland and Li, 2008, and others). The only exception to this trend of using non-F0 acoustic correlates for tone identity is Cantonese. The LF tone is often produced with some amount of glottalization, but this property has been shown to not function as a consistent perceptual cue for native Cantonese listeners (Vance, 1976). F0 is thought to possibly be the sole acoustic cue for Cantonese tone perception (Francis et al., 2008; see also Ciocca et al., 2002 and Lee et al., 2002).

In light of the above observations, it is possible that the Theory of Adaptive Dispersion could accurately predict cross-language tone-system tone-space and tonal dispersion trends if a model with multiple acoustic dimensions were created and tested. For instance, adding an amplitude variable to an lmer comparing Igbo and Mandarin might produce results consistent with the TAD prediction that the tone-space of the language with the larger tone inventory (Mandarin) > that of the language with the smaller inventory (Igbo). Similarly, adding a voice quality Variable to an lmer comparing Mandarin and Yoruba might potentially produce results consistent with the TAD prediction that, regarding degree of tone dispersion, Yoruba < Mandarin at onset and midpoint (since at onset and midpoint, Yoruba < Mandarin in tone-space size); and Yoruba = Mandarin at offglide (since at offglide, Yoruba = Mandarin in tone-space size).

Given that F0 is apparently the only cue to tone identity in Cantonese (Francis et al., 2008 and others), it would seem unlikely that adding a third (or fourth…) dimension to the tone-space and tone-dispersion models would affect outcomes involving comparisons with Cantonese. Indeed, the idea that Cantonese uses only F0 to differentiate its tones may explain why

Cantonese often displays both a greater overall tone space as well as greater degree of tone dispersion relative to languages with smaller tone inventories. Such extra expansion and dispersion, while not predicted by the TAD, may be necessary in order to ensure sufficient contrast between the tones of Cantonese (particularly at tone onset, and midpoint, where the Cantonese tones are especially crowded within its tone space). With regard to Cantonese, it is particularly curious that its M tone is so far from the H tone (see Figure 2.1). If the Cantonese M tone were even 1 ST higher across its trajectory, it would be more clearly differentiated from the LF, MR, and L tones at onset and midpoint. Likewise, the Cantonese MR tone overlaps the LR and L tones at onset, M tone at midpoint, and H tone at offglide; plus, its trajectory overlaps those of the L and LF tones until timepoint k5. If the MR tone started lower, rose more sharply, and ended higher, it would be more easily distinguished from the surrounding tones. With these observations in mind, it seems that Cantonese has a tendency to have comparatively low onsets and midpoints, and extra tonal spread at offglides. Relative to the H tone, the Cantonese L tone is lower at both tone onset and midpoint than that of Igbo. Compared to Yoruba, the Cantonese L tone is lower at onset and midpoint, and its M tone is lower at all three timepoints. Compared to Thai, the Cantonese R tone is lower at onset, and both its M and L tones are lower at onset and midpoint. Such results recall studies of cross-language vowel-category organization found that the location of similar vowels in acoustic vowel spaces differed across languages (e.g., Disner, 1983; Bradlow, 1995). Bradlow, for instance, found that the F2 of the English vowels [i, e, o, u] is systematically significantly higher than the F2 in Spanish vowels [i, e, o, u]. These results were accounted for by a language-specific base-of-articulation property: due to different bases of articulation across languages, sound categories that have the same phonological features and are located in similar positions in acoustic space across languages may actually have different phonetic realizations. A similar base-of-articulation property may be present in, and account for differences between, some of the tone-category location differences observed in the data of the current study. Specifically, Cantonese may have a different tonal base of articulation than some of the other languages. I speculate that as the number of tones in the inventory increases, languages may systematically alter the phonetic realization of their tones in order to enhance their auditory distinctiveness. One simple explanation is that, by systematically lowering tonal F0 at one or more points along the tonal trajectory, the language takes advantage of a greater portion of the frequency range to which listeners are most sensitive. It may also be possible that systematic lowering of tonal F0 may cause tones to be perceived in a more categorical (less continuous) manner, which would in turn make it easier for listeners to identify and discriminate the tones. This would be consistent with the notion that tonal category boundaries are determined by not only linguistic experience, but also regions of natural auditory sensitivity (see Francis, Ciocca, and Kei Chit Ng, 2003, for a discussion on this topic).

Finally, it is possible that the level-tone systems of Yoruba and Igbo may reasonably be compared with one another, but not with the contour-tone systems of Cantonese, Mandarin, and Thai. If so, it may be possible that level-tone system tone-spaces and tone dispersion could be accounted for by the TAD – even if mean F0 remains the only acoustic correlate under consideration. Recall that work on vowel systems indicated that most languages have the point vowels [a-i-u], and that other vowels are added to inventories around these three vowels. Figures 2.4-2.5 illustrate quite nicely that the H and L tones might be the point tones for level-tone languages, and that the M tone was simply added in Yoruba to that basic tone inventory.

Additionally, at tone midpoint and offglide at least, the tone spaces of Yoruba and Igbo are equivalently-sized, and the Yoruba M tone is well dispersed from its H and L tones. Further evidence for the notion that level-tone and contour-tone languages are not comparable comes from the fact that this study appears to indicate that both tone-space size may generally be determined first by the *type* of tones in the inventory. (I cannot comment on the possibility that tone dispersion is determined in this fashion, as I only compared tones of two languages in each of the *ToneDisp* models, and therefore lack sufficient evidence to support such a notion.) The results of the *ToneSpace* models suggest that at tone onset and midpoint, level-tone-only languages have smaller tone spaces than contour-tone languages. It is possible that contour-tone languages need more acoustic space to accommodate the full pitch-excursion needed for distinct tones at those timepoints. That said, the results of *ToneSpaceOffglide* indicated that all the languages but Thai had equivalently-sized tone spaces at offglide. These results may suggest that tone offglide F0 bears extra weight as an acoustic cue to tone identity for all the languages but Thai. The idea that tone offglide is special for level-tone identity in particular is supported by the literature. As mentioned in chapter one, despite the fact that Igbo tones are considered level, their phonetic values are actually determined according to their targets, found at the end of the timespan of the associated tone-bearing unit (Akinlabi and Liberman, 2000:5). Also recall that Hombert (1976) found that when Yoruba L tones were manipulated to have a level (as opposed to a falling) offglide, native listeners misidentified L-L sequences as L-M sequences and M-L sequences as M-M sequences. Tone offglide may be special for Mandarin and Cantonese tone identity as well. Chao (1968) suggests, for instance, that Mandarin tones converge gradually to a contour that seems to conform to purported underlying F0 targets. Li (2004) found that Cantonese listeners performed comparatively poorly on Cantonese-tone perception tasks unless the entire tone was presented. The exception of Thai among the other contour tone languages in *ToneSpaceOffglide* is not necessarily surprising if F0 direction is of greater importance than offglide F0 for Thai tone perception (Pike, 1948; Gandour, 1983).

5.4. Alternative analyses of cross-language tone-spaces

The current study is innovative in part because it defines the size of the tone space as the mean F0 distance (in ST) between each language's highest and lowest ("extreme") tones at equidistant timepoints across the tonal trajectory. However, this may not be the only – or the optimal – way to define the acoustic tone space. In this section, I compare the five languages' tone spaces in two alternative ways and discuss the results within the framework of the TAD. In section 5.4.1, the tone space is shown as plots of F0 offglide x F0 onset, following the method suggested by Barry and Blamey (2004). In section 5.4.2, tone space size is defined as the difference between the maximum and minimum raw F0 values produced across a small subset of the data (in the syllable [ba]).

Also, recall that the *ToneSpace* results suggest level-tone and contour-tone languages may organize their tone spaces in very different ways. In section 5.4.3, I investigate whether tone space size differs as a function of language type (contour vs. level). Tone space size is once again defined as the mean F0 distance (in ST) between the highest and lowest tones at the tonal onset, midpoint, and offglide. In the following sections, all data are taken from the same set as that used for earlier analyses.

5.4.1. Cross-language tone-spaces as plots of F0 offglide x F0 onset

As discussed in Barry and Blamey (2004), Gandour (1978) suggests that there are five acoustic dimensions that account for listeners' perceptual judgments about tone: (a) average pitch, (b) pitch direction, (c) length, (d) extreme endpoint, and (e) slope. Barry and Blamey compared Cantonese tone productions in normally-hearing adults, normally-hearing children, and cochlear-implanted children. Citation-form tones were elicited via a picture-naming task involving 15 presentations of each of the six tone types on various (unreported) syllables, for a total of 90 items per participant. The authors plotted the tone productions in an F0 offglide x F0 onset (Hz) space. They chose this method of analysis because it captures all the aforementioned dimensions except length. In particular, since F0 onset and offglide are the only points plotted, the method highlights (a) pitch level differences between tone types, and (b) pitch movement across the tone. Ellipses surrounding tokens of each tone illustrate within-category differences between those tokens. Figure 5.1, reproduced from Barry and Blamey (2004:1743), shows tone plots for two typical normally-hearing adult speakers of Cantonese. A1 is male and A2 is female. Note that the most differentiated of the six tones are H (55), MR (25), and LF (21). One way to define the periphery of the tone space is by the triangle that would result from connecting with lines the centers of these three tones' ellipses. Another way to define the periphery of the tone space is by the shape that results from connecting the centers of all six tones' ellipses. Also note the tonal crowding: M (33), LR (23), and L (22) are crowded within the space.

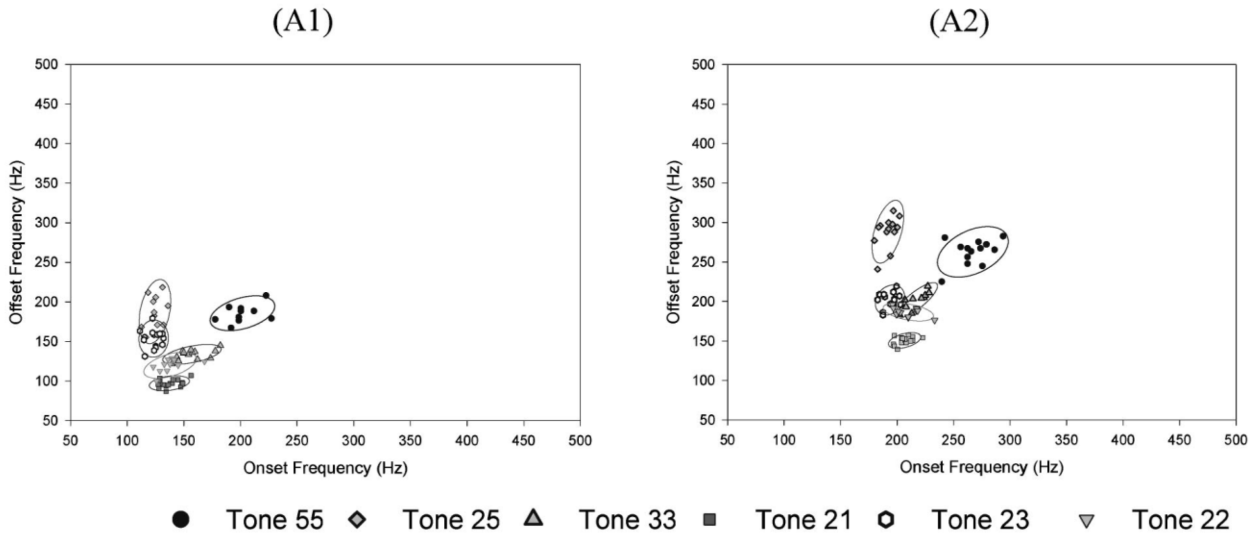
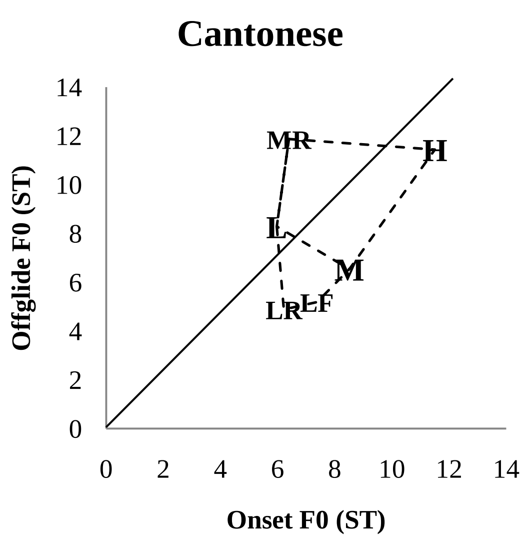
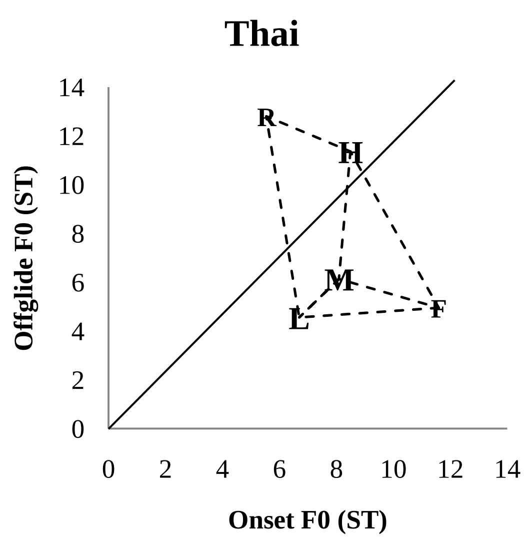


Figure 5.1. Tone plots of two adult Cantonese speakers from Barry and Blamey (2004)

The three groups of talkers under investigation in the Barry and Blamey study were clearly identifiable on observations of the locations of the F0 onset x F0 offglide points, and the degree of differentiation of the ellipses, within the tonal space. This approach to acoustic analysis of tone therefore enhanced understanding of tone production based on auditory analyses. Given its success in highlighting differences in tone productions across different populations of speakers of a single language, I surmise that the Barry and Blamey methodology might also be used to compare tones across languages. In these types of plots, points corresponding to level tones would be expected to fall about halfway between the two axes, if those level tones do indeed have roughly equivalent onset and offglide F0 values. Rising tones are expected to fall closer to the y-axis (lower onset, higher offglide), and falling tones are expected to cluster closer to the x-axis (higher onset, lower offglide).

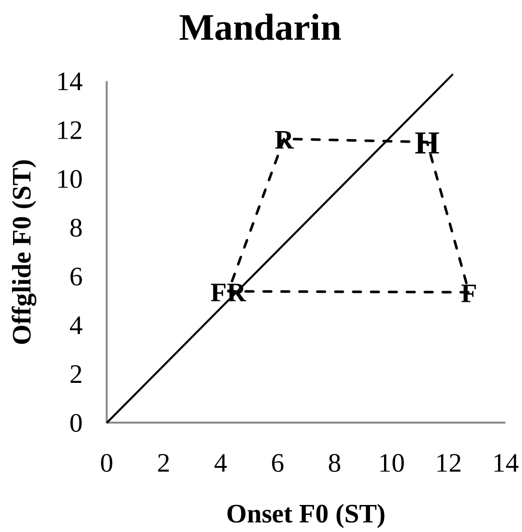
Figure 5.2 illustrates the Cantonese, Thai, Mandarin, Yoruba, and Igbo tone spaces as plots of F0 offglide (mean F0 at timepoint k9) x F0 onset (mean F0 at timepoint k1), in semitones. The accompanying tables also display the F0 at onset and offglide, and are reproduced from Figures 2.1-2.5. Points corresponding to level tones are in larger font, so as to be differentiable from points corresponding to contour tones. Dashed lines connecting the tone points span the F0 range used by talkers and define the extent of the tone spaces. The y=x diagonal is shown as well. These figures lack ellipses because they are plots of each tone's mean F0 across multiple variables (talkers, sex, items, etc.), as opposed to tokens of each tone produced by individual talkers.



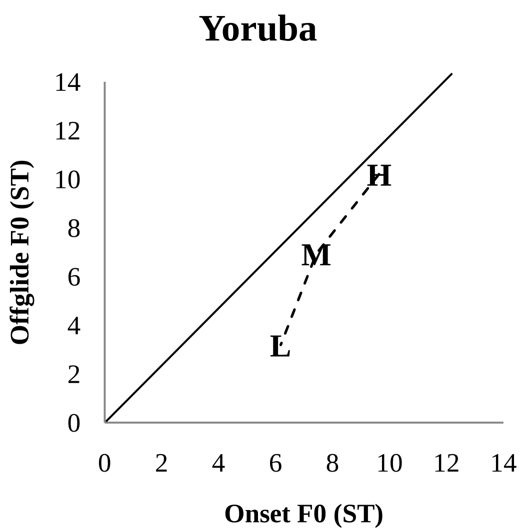


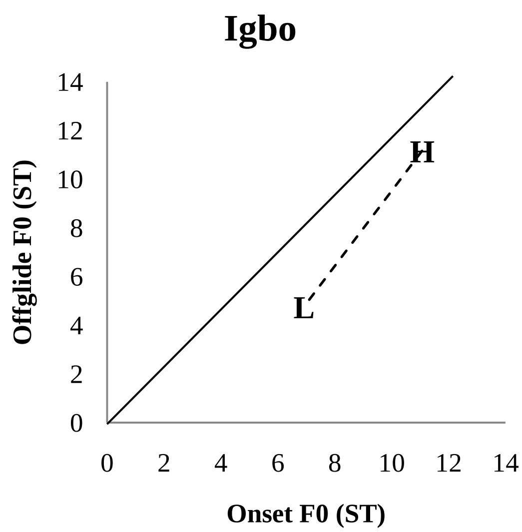
Cantonese		
Tone	**Onset F0 (ST)**	**Offglide F0 (ST)**
H	11.509	11.445
L	5.965	8.283
LF	7.381	5.189
LR	6.219	4.886
M	8.515	6.536
MR	6.397	11.868

Thai		
Tone	**Onset F0 (ST)**	**Offglide F0 (ST)**
H	8.496	11.357
R	5.549	12.804
F	11.601	4.955
L	6.700	4.552
M	8.098	6.137



Mandarin		
Tone	**Onset F0 (ST)**	**Offglide F0 (ST)**
H	11.249	11.505
R	6.230	11.636
FR	4.270	5.379
F	12.725	5.341





Yoruba		
Tone	**Onset F0 (ST)**	**Offglide F0 (ST)**
H	9.638	10.203
M	6.182	3.189
L	7.427	6.934

Igbo		
Tone	**Onset F0 (ST)**	**Offglide F0 (ST)**
H	11.071	11.160
L	6.926	4.760

Figure 5.2. Onset F0 x offglide F0 plots, and onset and offglide F0 values, for Cantonese, Thai, Mandarin, Yoruba, and Igbo

As expected from re-examination of figures 2.1-2.5, most level tones are located alongside the y=x diagonal; none fall precisely on the diagonal, because none are precisely level. The Mandarin FR tone is also located at the y=x diagonal, illustrating that its onset and offglide F0 values are nearly equivalent. Also as expected, most rising tones (e.g., the Cantonese MR and Thai and Mandarin R tones) are located in the top left quadrant of the space, as their onset F0 values are lower than their offset F0 values. Likewise, falling tones (e.g., the Cantonese LF and Thai and Mandarin F tones) are located in the lower right quadrant of the space, as their onset F0 values are higher than their offglide F0 values. Note also that the Igbo, Yoruba, and Thai L tones are located below the diagonal, illustrating that their F0s are lower at offglide than at onset, while the Cantonese L tone is above the diagonal, indicating that its F0 is higher at offglide than at onset. Additionally, the Cantonese, Thai, and Yoruba M tones fall below the diagonal, as they drop in pitch across their trajectories. Also, the triangular shape of the Cantonese space echoes that of the Barry and Blamey Cantonese space. (That said, the L and LR tones in the current study are located in different places in the tone space than those of the Barry and Blamey paper. Recall from chapter two that the Cantonese tonal F0 values in the current study differ from those reported elsewhere in the literature, so this is unsurprising.)

The above plots provide some interesting insights into cross-language tone-system structures. Overall, each of the languages disperses its tone categories across the onset F0 x offglide F0 space, indicating that F0 trajectory is indeed a key acoustic correlate used to distinguish the tones of these languages. This may be taken as evidence that the tone spaces of these languages may in fact be reasonably defined by F0 trajectory alone. Importantly, this also provides support for the TAD hypothesis that sound categories will be well-dispersed across the acoustic space and will thereby be highly contrastive (save for the Cantonese LR and LF tones, which overlap to a considerable degree in this space). This also rectifies the mystery of how contour tones could possibly be perceptually contrastive if their F0 values overlap at any point in their trajectories. If listeners attend to *both* tonal onset and offglide F0 values, even tones whose F0 values overlap become quite differentiable. For instance, the high degree of tonal crowding observed at onset and offglide in Cantonese (figure 2.1) may not negatively impact perception if listeners reconcile each tone's onset vs offglide pitch-height difference. These results do *not* appear to support the TAD notion that languages with larger tone inventories will have expanded tone spaces relative to those with smaller inventories. In light of the above discussion, however, this is not wholly surprising. If the tones are differentiable by their onset x offglide F0 values, expansion of the overall tone space area might well be unnecessary and redundant and therefore inefficient.

5.4.2. Cross-language tone-spaces as max – min F0 in token syllable [ba]

In this section I define the tone-space periphery according to the extremes of the pitch range employed during speech (specifically, during tone production). Tone space size is defined as the difference between the maximum and minimum raw F0 values produced across a small subset of the data (in the syllable [ba]). The upper bound of the tone space is therefore the single highest F0 value across all productions of [ba] by speakers of that language. Likewise, the lower bound of each language's tone space is the single lowest F0 value across all [ba] productions. The syllable [ba] was chosen because, out of all the syllables in the data set, it most often is a meaningful word when produced with the languages' tones.

There are multiple benefits to defining the periphery of the tone space this way. First, recall that languages with two-tone inventories (e.g., Igbo) would be excluded from tonal dispersion analyses if its highest and lowest tones defined the space, because it would be a confound to consider those tones to also be located *within* the tone space. By defining the tone space according to raw F0 extremes produced during speech, all tones other than those delineating the edges of the tone space are considered to fall within the tone space and may therefore be included in tests of degree of tonal dispersion within the space. (Further analyses of tone dispersion are not performed here, but are left for future work.) Additionally, the acoustic space is constrained by F0 values produceable (and indeed produced) by the human vocal tract during natural speech, a key tenet of the TAD (see chapter one). Additionally, this provides a realistic view of the tonal pitch range not afforded by other possible methods. For instance, the pitch range could be defined by "vocalese" exercises, in which the participant vocalizes as high and low as possible, but many people (in particular, trained vocalists) can easily exceed their natural speech pitch range during vocalese exercises.

Figure 5.3 illustrates the languages' tone spaces, defined as the maximum and minimum F0 values produced in utterances of the syllable [ba]:

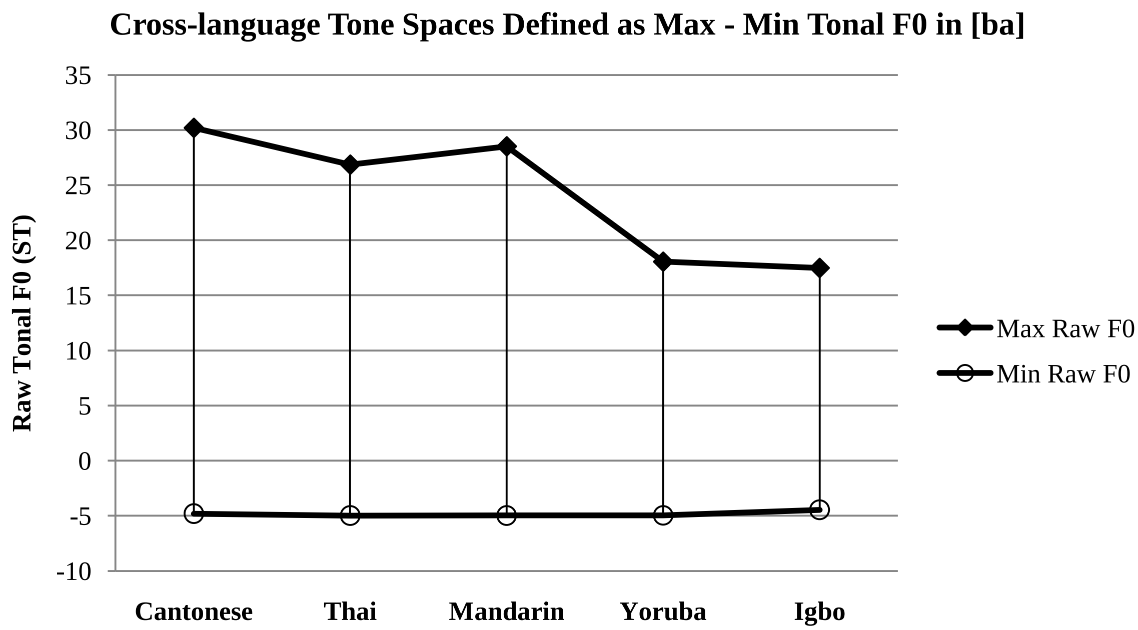


Figure 5.3. Maximum and minimum tonal F0 in productions of the syllable [ba]

	Cantonese	Thai	Mandarin	Yoruba	Igbo
Max Raw F0	30.200	26.864	28.526	18.056	17.488
Min Raw F0	-4.805	-4.980	-4.973	-4.958	-4.458
Max - Min F0	35.005	31.844	33.499	23.014	21.946
Max Raw F0	30.200	26.864	28.526	18.056	17.488

Table 5.2. Maximum and minimum F0 values in productions of the syllable [ba]

Like the *ToneSpace* analyses conducted in chapter four, these results do not provide clear support for the TAD hypothesis that a language with a larger tone inventory will have an expanded tone space relative to a language with a smaller tone inventory. Upon visual inspection of these data, it appears that Cantonese, with the largest tone inventory (6 tones), also has the largest tone space. However, the tone space of Thai, with 5 tones, is smaller than that of Mandarin, with 4 tones. In fact, the tone spaces of Mandarin and Cantonese are very similar in size; the Mandarin tone space is only 1.67 ST smaller than that of Cantonese. The Cantonese, Mandarin, and Thai tone spaces are much larger than that of Yoruba (3 tones). Yet, the Igbo and Yoruba tone spaces are, in effect, equivalently sized – that of Igbo is only 0.568 ST smaller than that of Yoruba.

These results do provide support for the notion that tone space size may differ as a function of tone-language type (level vs. contour): the tone spaces of the contour-tone languages are all much larger than those of the level-tone languages. This possibility is further explored in the following section.

5.4.3. Tone-space size as a function of language type

The following three models – *LangTypeToneSpaceOnset*, *LangTypeToneSpaceMidpoint*, and *LangTypeToneSpaceOffglide* – examine whether tone-space size differs as a function of language type (contour vs. level). Like the chapter four *ToneSpace* models, they compare at onset, midpoint, and offglide the F0 distances between the languages' highest (top) and lowest (bottom) tonal F0 values. However, these models compare just two tone spaces: that of the three contour-tone languages combined vs. that of the two level-tone languages combined. All values are significant at $p \leq 0.05$ (are not corrected), because each analysis contains just one pairwise comparison.

5.4.3.1. *LangTypeToneSpaceOnset*

The *LangTypeToneSpaceOnset* models compare the F0 difference (in ST) between the highest (top) and lowest (bottom) tones of the languages at tonal onset. Figure 5.4 shows the top-bottom tone mean F0 at timepoint k1 for the two language types. Each data point has standard error bars.

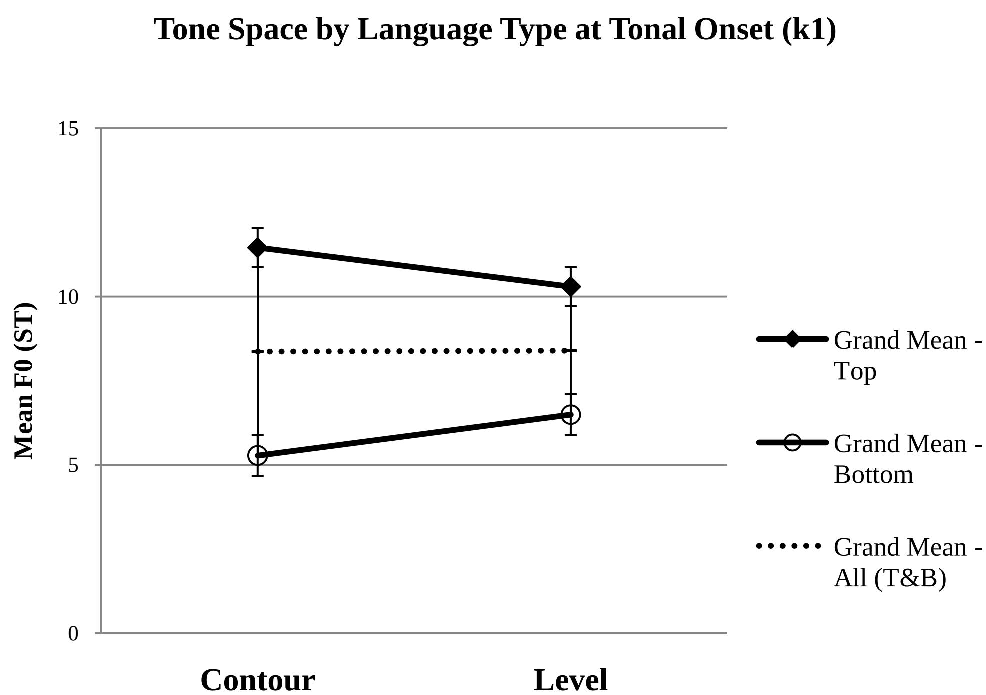


Figure 5.4. Tone-space size across the two language types at tonal onset

Item	Contour	Level
Grand Mean - Top	11.449	10.291
Grand Mean - Bottom	5.275	6.490
Grand Mean – All (T&B)	8.362	8.391
Grand Mean T-B	6.174	3.801

Table 5.3. Tone-space size F0 (ST) values across the two language types at tonal onset

Observe that the Grand Mean – All (T&B) values are nearly the same across the two language types. In addition, note that the differences between the language types' top vs. bottom tone

Grand Mean F0s differ by nearly 2.4 ST. Table 5.4 shows the fixed-effects results of the *LangTypeToneSpaceOnset* lmer.

***LangTypeToneSpaceOnset:* Contour vs. Level**				
	Est	**St.E**	**t-val**	**pMCMC**
LangTypeLevel	-1.224	1.763	-0.69	0.0722
ToneB	-6.168	0.089	-69.3	**0.0001**
LangTypeLevel:ToneB	2.398	0.095	25.16	**0.0001**

Table 5.4. Summary of the results of the *LangTypeToneSpaceOnset* lmer

The results of the *LangTypeToneSpaceOnset* model are summarized below:

1. <u>There is no main effect of language type.</u> The Grand Mean F0 did not differ as a function of language type.
2. <u>Overall, there is a significant difference between the top and bottom tones.</u> The bottom tone was about 6 ST lower on average than the top tone. This indicates that the top and bottom tones are well-differentiated overall.
3. <u>The interaction of tone and language type is significant.</u>

The results of *LangTypeToneSpaceOnset* corroborate the observations of the data in Figure 5.4: tone-space size at onset indeed differs as a function of language type; specifically, contour-tone languages have a larger tone space than level-tone languages at onset.

5.4.3.2. *LangTypeToneSpaceMidpoint*

The *LangTypeToneSpaceMidpoint* models compare the F0 difference (in ST) between the highest (top) and lowest (bottom) tones of the languages at tonal midpoint. Figure 5.5 shows the top-bottom tone mean F0 at timepoint k5 for the two language types. Each data point has standard error bars.

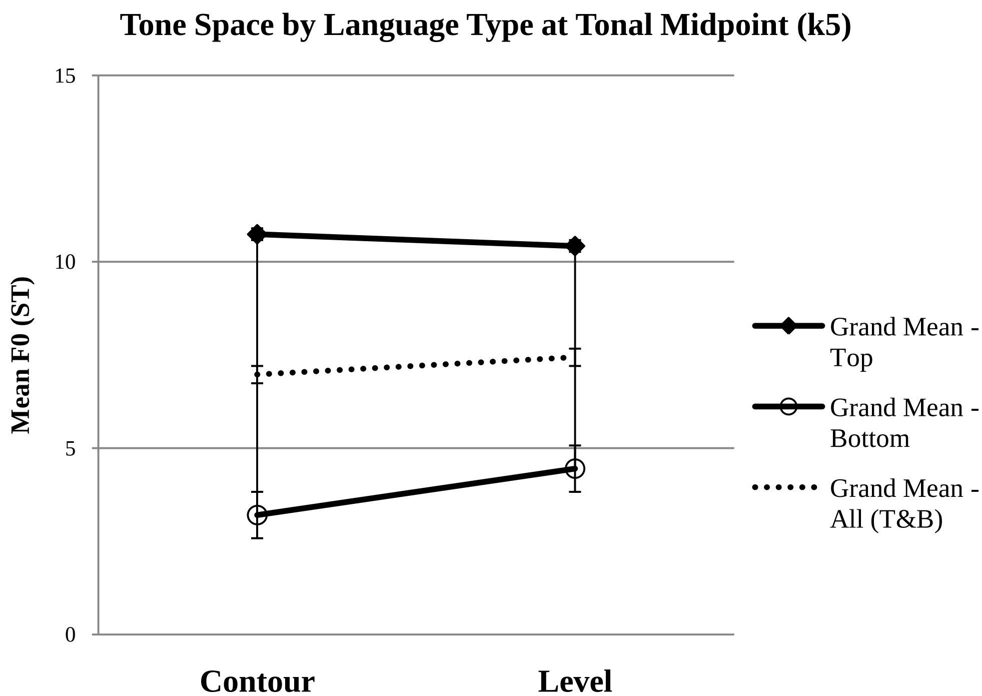


Figure 5.5. Tone-space size across the two language types at tonal midpoint

Item	Contour	Level
Grand Mean - Top	10.737	10.422
Grand Mean - Bottom	3.204	4.448
Grand Mean – All (T&B)	6.971	7.435
Grand Mean T-B	7.533	5.974

Table 5.5. Tone-space size F0 (ST) values across the two language types at tonal midpoint

Observe that the two language types' Grand Mean – All (T&B) values differ only by 0.46 ST. In addition, note that the differences between the language types' top vs. bottom tone Grand Mean

F0s differ by about 1.6 ST. Table 5.6 shows the fixed-effects results of the *LangTypeToneSpaceMidpoint* lmer.

***LangTypeToneSpaceMidpoint:* Contour vs. Level**				
	Est	**St.E**	**t-val**	**pMCMC**
LangTypeLevel	-0.376	1.605	-0.230	0.6812
ToneB	-7.534	0.122	-61.810	**0.0001**
LangTypeLevel:ToneB	1.572	0.145	10.810	**0.0001**

Table 5.6. Summary of the results of the *LangTypeToneSpaceMidpoint* lmer

The results of the *LangTypeToneSpaceMidpoint* model are summarized below:

1. There is no main effect of language type. The Grand Mean F0 did not differ as a function of language type.
2. Overall, there is a significant difference between the top and bottom tones. The bottom tone was about 7.5 ST lower on average than the top tone. This indicates that the top and bottom tones are well-differentiated overall.
3. The interaction of tone and language type is significant.

The results of *LangTypeToneSpaceMidpoint* corroborate the observations of the data in Figure 5.5 and echo those of *LangTypeToneSpaceOnset*: tone-space size at midpoint differs as a function of language type; contour-tone languages have a larger tone space than level-tone languages at midpoint.

5.4.3.3. *LangTypeToneSpaceOffglide*

The *LangTypeToneSpaceOffglide* models compare the F0 difference (in ST) between the highest (top) and lowest (bottom) tones of the languages at tonal offglide. Figure 5.6 shows the top-bottom tone mean F0 at timepoint k9 for the two language types. Each data point has standard error bars.

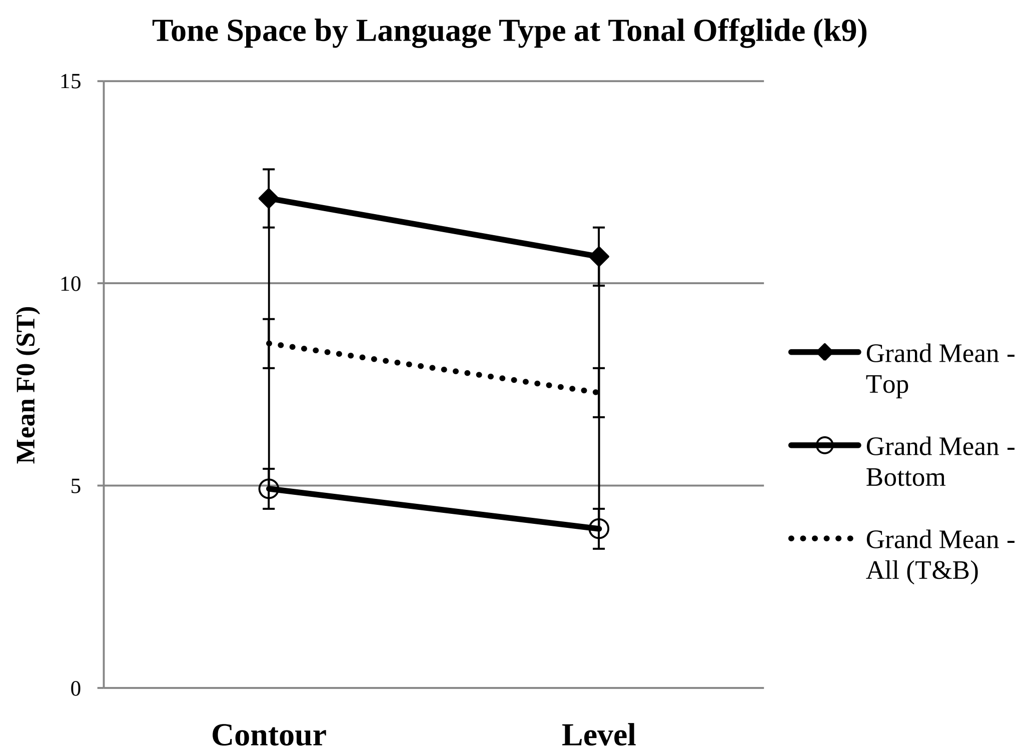


Figure 5.6. Tone-space size across the two language types at tonal offglide

Item	**Contour**	**Level**
Grand Mean - Top	12.101	10.663
Grand Mean - Bottom	4.924	3.936
Grand Mean – All (T&B)	8.513	7.300
Grand Mean T-B	7.177	6.727

Table 5.7. Tone-space size F0 (ST) values across the two language types at tonal offglide

Observe that the two language types' Grand Mean – All (T&B) values only differ by about 1.2 ST. In addition, note that the differences between the language types' highest-tone vs. lowest-tone Grand Mean F0s differ by only 0.45 ST. Table 5.8 shows the fixed-effects results of the *LangTypeToneSpaceOffglide* lmer.

LangTypeToneSpaceOffglide: **Contour vs. Level**				
	Est	**St.E**	**t-val**	**pMCMC**
LangTypeLevel	-1.480	1.573	-0.940	0.216
ToneB	-7.189	0.158	-45.570	**0.0001**
LangTypeLevel:ToneB	0.456	0.250	1.820	0.067

Table 5.8. Summary of the results of the *LangTypeToneSpaceOffglide* lmer

The results of the *LangTypeToneSpaceOffglide* model are summarized below:

1. There is no main effect of language type. The Grand Mean F0 did not differ as a function of language type.
2. Overall, there is a significant difference between the top and bottom tones. The bottom tone was about 7.2 ST lower on average than the top tone. This indicates that the top and bottom tones are well-differentiated overall.
3. The interaction of tone and language type is not significant.

The results of *LangTypeToneSpaceOffglide* indicate that tone-space size at offglide does not differ as a function of language type; contour-tone languages and level-tone languages appear to have equivalently-sized tone spaces at offglide.

5.4.3.4. Summary of *LangTypeToneSpace* analyses

The flowchart in Figure 5.7 illustrates the results of the *LangTypeToneSpace* analyses.

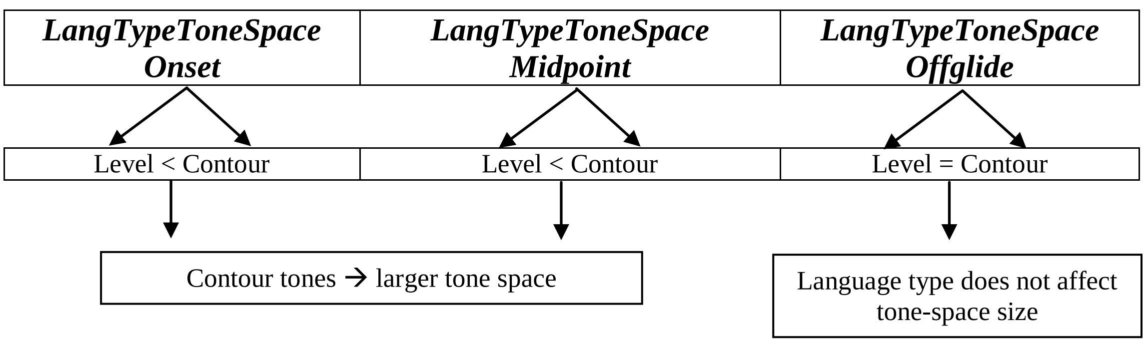


Figure 5.7. Flowchart summarizing the *LangTypeToneSpace* analyses

As illustrated in the flowchart, there does appear to be a significant effect on tone space size of tone language type, but only at tonal onset and midpoint. At these two timepoints, the level-tone-language space is smaller than the contour-tone space. This echoes the trend observed in section 5.4.2, where the tone space was defined as the difference between the single highest and single lowest F0 produced in the syllable [ba]. It also echoes the results of *ToneSpaceOnset* and, for the most part, *ToneSpaceMidpoint* (recall that the tone space of Thai was equivalent in size to those of Yoruba and Igbo in *ToneSpaceMidpoint*). Because both Igbo and Yoruba have fewer tones than any of the contour-tone languages, the results of the *LangTypeToneSpaceOnset* and

Midpoint analyses also appear to support the TAD hypothesis that languages with larger tone inventories will have larger tone spaces, relative to languages with smaller tone inventories.

On the other hand, the tone spaces of the two language types were equivalently sized at offglide. The results of *LangTypeToneSpaceOffglide* thus support the hypothesis that the tone space will be fixed in size, regardless of the size of the tonal inventory. The results of this analysis are roughly consistent with the results of *ToneSpaceOffglide*, in which the tone spaces of the level-tone languages were found to be equivalent to those of Mandarin and Cantonese (but not Thai). However, there is a notable difference between the two language types' tone spaces at offglide: the level-tone-language speakers utilized a lower overall pitch range than the contour-tone-language speakers. It is possible that the level-tone languages have a tendency to have comparatively lower offglides, thus still making them qualitatively different from the contour-tone languages at offglide. All told, it could be argued that there *is* a significant effect of language-type on the acoustic tone space, whether it be the tone-space size or the tone-space pitch range.

5.5. Indications for further research

In this section I outline various experiments that would help to clarify some of the issues raised in this dissertation.

One of the most obvious follow-up studies would construct models of tone-systems that take into account other variables, such as phonation type; tone duration; talker sex and age; etc. The models reported in this study serve to capture the overall trends of tone-system organization across languages; they may well miss certain subtleties that could come to light with methodical inclusion of other variables. For instance, it is very possible that some populations of talkers, e.g., males and females, might display differently-sized tone spaces, and that the conclusions reached in the current study might more accurately describe one population over another. (Despite its benefits and utility for this study, the ST scale does not normalize for pitch range, so nuances in tone-spaces due to sex and/or other inter-talker pitch-range differences may be missed in these analyses). Additionally, accounting for phonation-type could be informative because phonation type is an additional cue to tone identity in, e.g., Mandarin (the FR tone is typically produced with creaky voice; see Chao, 1948 and many others).

In general, comparisons of the tone systems of more languages would also help to test whether the TAD can accurately predict cross-language tone-system acoustics. For instance, Southern Vietnamese has 5 tones that are distinguished primarily by F0, and Northern Vietnamese has 6 tones that are distinguished by F0 and voice-quality characteristics (Kirby, 2010). Both dialects have tones that could be considered level, or at least simple (as opposed to complex): both have a relatively level high tone and a mid tone that falls about 50 Hz across its trajectory. Both also have, e.g., complex falling-rising (dipping) contour tones. However, Northern Vietnamese appears to have a low falling tone that the southern dialect seems to lack. Considering the similarities and differences between the two, as well as the similarities and differences between other languages with the same number of tones (e.g., Thai, with 5 tones, or Cantonese, with 6 tones), adding such languages to future investigations could clarify the extent to which the various results in this study are generalizable to other languages, and/or the extent to which these findings are language-specific. It is possible, for instance, that Thai and Southern Vietnamese would have similarly-sized tone-spaces, degrees of tone crowdedness, and tone-category locations. It is also possible that these two languages would differ in a significant way along one or more of these dimensions. As such, such experiments could help to illuminate intricacies of the theory that tone-space-size, degree of tone crowdedness, and location of tone categories within the tone space are determined first by the type, then the number, of tones.

Additionally, a set of studies are needed to examine whether the conclusions reached in this thesis extend to perception of tone contrasts. Recall that the hypotheses of the TAD that were tested here are based on the idea that tones are organized in acoustic space in such a way as to make them maximally contrastive for the listener. Follow-up studies would be indicated to test whether, e.g., tones of languages that have more crowded tone spaces (Mandarin, Cantonese) are harder for listeners to distinguish than tones of languages that have less crowded tone spaces (Igbo, for one). Other experiments are indicated to investigate whether the tones of languages with larger overall tone spaces (generally speaking, those of the contour-tone languages) are more easily distinguished than the tones of languages with smaller overall tone spaces (generally speaking, those of the level-tone-only languages).

Another logical test of the robustness of the findings reported herein is to examine and compare these tone systems using tones excised from a carrier-sentence context. As discussed in chapter one, each of the languages examined in this study are subject to tone-alternation rules. For instance, two adjacent falling-rising (dipping) Mandarin tones are subject to sandhi, wherein the first of the two changes to a rising tone. Igbo and Yoruba are subject to, e.g., downdrift and tone spreading, other processes that affect the phonetic realization of the tones. Careful construction of appropriate carrier sentences, serving to control (to the degree possible) tone alternations, could provide insight as to tone-space organization in more natural speech.

REFERENCES

Abramson, A. (1962). The vowels and tones of Standard Thai: Acoustical measurements and experiments. *Indiana University Research Center in Anthropology, Folklore, and Linguistics, 20*, Bloomington: University of Indiana.

Abramson, A. (1972). *Tonal experiments with whispered Thai.* In A. Valdman (Ed.), *Papers in linguistics and phonetics to the memory of Pierre Delattre* (pp. 29-55). The Hague: Mouton.

Abramson, A. (1975). *The tones of Central Thai: Some perceptual experiments*. In J. G. Harris & J. Chamberlain (Eds.), *Studies in Tai Linguistics* (pp 1-16). Bangkok: Central Institute of English Language.

Abramson, A. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, *23,* 319-325.

Abramson, A. (1979). The coarticulation of tones: an acoustic study of Thai. In T. Thongkum, V. Panupong, P. Kullavanijaya, & M.R.K. Tingsabadh (Eds.), *Studies in Tai and Mon-Khmer Phonetics and Phonology in Honor of Eugénie J.A. Henderson* (pp. 1-9). Bangkok: Chulalongkorn University Press.

Akinlabi, A. (1985). *Underspecification of tones in Yoruba*. PhD dissertation, University of Ibadan, Ibadan, Nigeria.

Akinlabi, A. & Liberman, A. M. (2000). In M.-J. Kim & U. Strauss (Eds.), *Proceedings of the North East Linguistics Society (NELS) 31*; Georgetown University, Washington, D.C., October 6-8, 2000. Amherst: GSLA Publications.

Al-Tamimi, J.-E. & Ferragne, E. (2005). Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. In *Proceedings of Interspeech 2005-Eurospeech, 9th European Conference on Speech Communication and Technology* (pp. 2465-2468); Lisbon, Portugal, September 4-8, 2005.

Anderson, L. (1959). Ticuna vowels, with special regard to the system of five tonemes. *Serie Linguistica Especial*, *1*, 76-119.

Asongwed, T. & Hyman, L.M. (1976). Morphotonology of the Ngamambo noun. In Hyman, L.M. (Ed.), *Studies in Bantu Tonology* (pp. 23-56).

Baayen, R.H. (2009). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R.* Cambridge: Cambridge University Press.

Barker, M. E. (1966). Vietnamese-Muong tone correspondences. In N.H. Zide & M.E. Barker (Eds.), *Studies in Comparative Austroasiatic Linguistics* (pp. 9-25). The Hague: Mouton.

Barry, J. & Blamey, P. (2004). The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese. *Journal of the Acoustical Society of America, 116*(3), 1739-1748.

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

Béarth, T. & Zemp, H. (1967). The phonology of Dan (Santa). *Journal of African Languages, 6*, 9-29.

Bergman, R. (1971). Vowel sandhi and word division in Igede. *The Journal of West African Languages, 8*, 13-25.

Bernstein, J. (1976). Vocoid psychoacoustics, articulation, and vowel phonology. *Natural Language Studies*. Ann Arbor: University of Michigan.

Bernstein Ratner, N. (1984). Patterns of vowel modifications in mother-child speech. *Journal of Child Language, 11*, 557-578.

Blicher, D.L., Diehl, R.L., & Cohen, L.B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics, 18*, 1-13.

Boersma, P., & Weenink, D. (2009). *Praat: doing phonetics by computer (Version 5.1.20)*. Retrieved November 18, 2009, from http://www.praat.org/.

Bradlow, A. R. (1993). *Language-Specific and Universal Aspects of Vowel Production and Perception: A Cross-Linguistic Study of Vowel Inventories*. PhD Dissertation, Department of Modern Languages and Linguistics, Cornell University, Ithaca, NY.

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America, 97*, 1916-1924.

Carrell, P. L. (1970). *A transformational grammar of Igbo.* Cambridge: Cambridge University Press.

Chan, C. L. (2009). *Triggerwave, an alignment tool.* Retrieved November 1, 2009 from http://groups.linguistics.northwestern.edu/documentation/triggerwave_home.html.

Chan, C.L. (2010). *NUAligner*. Retrieved from http://groups.linguistics.northwestern.edu/documentation/nualigner_home.html

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

Chang, K. (1953). On the tone system of the Miao-Yao languages. *Language, 29,* 374-378.

Chao, Y. R. (1948). *Mandarin Primer*. Cambridge: Harvard University Press.

Chao, Y.R. (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Cheng, C. C. (1973). A quantitative study of Chinese tones. *Journal of Chinese Linguistics, 1*, 93-110.

Chuang, C.-K. & Hiki, S. (1972). Acoustical features and perceptual cues of the four tones of standard colloquial Chinese. *Journal of the Acoustical Society of America*, 22(1A), 146.

Ciocca, V., Francis, A. L., Aisha, R., & Wong, L. (2002). The perception of Cantonese lexical tones by early-deafened cochlear implantees. *Journal of the Acoustical Society of America, 111*(5), 2250-2256.

Clark, M.M. (1990). *The tonal system of Igbo*. Dordrecht: Foris Publications.

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

Cope, A.T. (1959). Zulu tonology. *Afrika und Ubersee, 43*(3), 190-200.

Crothers, J. (1978). Typology and universals of vowel systems. In J.H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of Human Language, Volume 2: Phonology* (pp. 93-152). Palo Alto: Stanford University Press.

das Gupta, K. (1968). *An introduction to Central Monpa*. Shillong, North-East Frontier Agency.

de Jong, K. & Obeng, S. G. (2000). Labio-Palatalization in Twi: Contrastive, Quantal, and Organizational Factors Producing an Uncommon Sound. *Language, 76*(3), 682-703.

Diehl, R.L, Lindblom, B., Hoemeke, K., & Fahey, R. (1996). On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics, 24*, 187-208.

Disner, S.F. (1983). Vowel quality: The relation between universal and language specific factors. *UCLA Working Papers in Phonetics, 58*.

Disner, S.F. (1984). Insights on vowel spacing. In Maddieson, I. (Ed.), *Patterns of Sounds* (pp. 136-155). Cambridge: Cambridge University Press.

Dreher, J. J. & Lee, P. C. (1966). Instrumental investigation of simple and paired Mandarin tonemes. *DARL Research Communication* (p. 13). Huntington Beach, CA: Advanced Research Laboratory, Douglas Aircraft Company.

Engstrand, O. & Krull, D. (1994). Durational Correlates of Quantity in Swedish, Finnish, and Estonian: Cross-language Evidence for a Theory of Adaptive Dispersion. *Phonetica, 51*, 80-91.

Erickson, D. (1974). Fundamental frequency contours of the tones of Standard Thai. *Pasaa, 4*, 1-25.

Ewan, W. (1975). Explaining the intrinsic pitch of vowels. *Journal of the Acoustical Society of America, 58*(S1), S40.

Fant, G. (1973). Stops in CV-syllables. In G. Fant (Ed.), *Speech Sounds and Features* (pp. 110-139). Cambridge: MIT Press.

Flege, J. E. (1989). Differences in inventory size affect the location but not the precision of tongue positioning in vowel production. *Language and Speech, 32*(2), 123-147.

Fok Chan, Y. Y. (1974). *A perceptual study of tones in Cantonese*. Hong Kong: University of Hong Kong Press.

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

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

Fromkin, V. A. (Ed.) (1978). *Tone: A Linguistic Survey*. New York: Academic Press.

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

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

Gandour, J.T. (1974). On the representation of tone in Siamese. In J. G. Harris & J. Chamberlain (Eds.), *Studies in Tai Linguistics* (pp. 170-195). Bangkok: Central Institute of English Language.

Gandour, J.T. (1978). The perception of tone. In V.A. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 41-76). New York: Academic Press.

Gandour, J.T. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics, 9*, 20-36.

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

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

Gandour, J. & Harshman, R. (1978). Cross-language differences in tone perception: A multi-dimensional scaling investigation. *Language and Speech, 21*(1): 1-33.

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

Gedney, W. (1965). Yay, a Northern Tai language in North Vietnam. *Lingua, 14*, 180-193.

Gendrot, C. & Adda-Decker, M. (2007). Impact of duration and vowel inventory size on formant values of oral vowels: an automated formant analysis from eight languages. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*; August 6-10, 2007, Saarbrücken, Germany.

Goldsmith, J. (1976). *Autosegmental phonology*. PhD dissertation, Department of Linguistics, Massachusetts Institute of Technology, Cambridge, MA.

Goldsmith, J. (1990). *Autosegmental and metrical phonology*. Oxford: Basil Blackwell.

Goldstein, L. (1976). *Perception of Siswati clicks*, UCLA.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Ho, A.T. (1976). The acoustic variation of Mandarin tones. *Phonetica, 33*, 353-367.

Hockett, C.F. (1955). A Manual of Phonology. *International Journal of American Linguistics*, *21*(4), Bloomington: Indiana University Publications in Anthropology and Linguistics.

Hoijer, H. (1945). Navaho phonology. *University of New Mexico Publications in Anthropology*.

Hombert, J.-M. (1976). Perception of tones of bisyllabic nouns in Yoruba. *SAL Supplement 6*, 109-122.

Hombert, J.-M., Ohala, J.J., & Ewan, W.G. (1979). Phonetic explanations for the development of tones. *Language, 55,* 37–58.

Howie, J. M. (1976). On the domain of tone in Mandarin: Some acoustical evidence. *Phonetica, 30,* 129-148.

Hyman, L.M. (1978). Historical Tonology. In V.A. Fromkin (Ed.) *Tone: A Linguistic Survey* (pp. 257-270). New York: Academic Press.

Hyman, L.M. (2001). Tone systems. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals: An international handbook* (pp. 1367-1380). Berlin: Walter de Gruyter.

Hyman, L.M. (2007). Universals of tone rules: 30 years later. In C. Gussenhoven & T. Riad (Eds.), *Tones and Tunes: Typological Studies in Word and Sentence Prosody* (pp. 1-34). Berlin: Mouton.

Hyman, L.M. & Schuh, R.G. (1974). Universals of tone rules: evidence from West Africa. *Linguistic Inquiry*, *5*, 81-115.

International Phonetic Association. (1999). *Handbook of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Jongman, A., Fourakis, M., & Sereno, J.A. (1989). The acoustic vowel space of Modern Greek and German. *Language and Speech*, *32*(3), 221-248.

Jongman, A., and Moore, C. (2000). The role of language experience in speaker and rate normalization processes. In *Proceedings of the 6th International Conference on Spoken Language Processing*, *I*, 62-65.

Kenstowicz, M. & Kisseberth, C. (1990). Chizigula tonology: the word and beyond. In S. Inkelas & D. Zec (Eds.), *The phonology-syntax connection* (pp. 163-194). Chicago: Center for the Study of Language and Information.

Khouw, E. & Ciocca, V. (2006). Acoustic and Perceptual Study of Cantonese Tones Produced by Profoundly Hearing-Impaired Adolescents. *Ear and Hearing*, *27*, 243–255.

Kirby, J. (2010). Dialect experience in Vietnamese tone perception. *Journal of the Acoustical Society of America*, *127*(6), 3749-3757.

La Velle, C. R. (1974). An experimental study of Yoruba tone. *Studies in African Linguistics, Suppl. 5,* 185-194.

Leben, W.R. (1971). On the segmental nature of tone in Thai. *Quarterly Progress Report, Research Laboratory of Electronics*, 221-224.

Leben, W.R. (1973). *Suprasegmental phonology*. PhD dissertation, Department of Linguistics, Massachusetts Institute of Technology, Cambridge, MA.

Lee, K.Y., van Hasselt, C.A., Chiu, S.N., & Cheung, D.M. (2002). Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children. *International Journal of Pediatric Otorhinolaryngology, 63*(2), 137-147.

Li, F.-K. (1965). The Tai and Kam-Sui languages. *Lingua*, *14*, 148-179.

Li, P-Y. (2004). Perceptual Analysis of the Six Contrastive Tones in Cantonese. In *Proceedings of the International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*; Beijing, China, March 28-31, 2004.

Liang, J. & van Heuven, V.J. (2004). Evidence for separate tonal and segmental tiers in the lexical specification of words: A case study of a brain-damaged Chinese speaker. *Brain and Language, 91*, 282-293.

Liberman, A.M., Schultz, J.M., Hong, S., & Okeke, V. (1993). The phonetic interpretation of tone in Igbo. *Phonetica, 50*(3), 147-160.

Liennes, M. (2003). *collect_pitch_data_from_files Praat script (Version 4.0.23)*. Retrieved November 18, 2009 from http://www.helsinki.fi/~lennes/praat-scripts/.

Liljencrants, J. & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language, 48*, 839-862.

Lin, M.C. (1965). The pitch indicator and the pitch characteristics of tones in Standard Chinese. *Acta Acoustica (China), 2*, 8-15.

Lindau, M. & Wood, P. (1977). Acoustic vowel spaces. *UCLA Working Papers in Phonetics, 38*, 41-48.

Lindblom, B. (1975). *Experiments in sound structure*. Plenary address, 8th International Congress of Phonetic Sciences; Leeds, England.

Lindblom, B. (1986). Phonetic Universals in Vowel Systems. In J.J. Ohala & J.J. Jaeger (Eds.), *Experimental Phonology* (pp. 13-44). Orlando: Academic Press.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modeling* (pp. 403-439). Dordrecht: Kluwer.

Lindblom, B. & Sundberg, J. (1969). A quantitative theory of cardinal vowels and the teaching of pronunciation. *STL-QPSR*, *10*(2-3), 19-25.

Lindblom, B. & Sundberg, J. (1971). A quantitative theory of cardinal vowels and the teaching of pronunciation. In T. Perren (Ed.), *Applications of linguistics*. Cambridge: Cambridge University Press.

Liu, S. & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech*, *47*, 109-138.

Livijn, P. (2000). Acoustic distribution of vowels in differently sized inventories – hot spots or adaptive dispersion? *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PERILUS) 11*.

Maddieson, I. (1972). Tone system typology and distinctive features. In *Proceedings of the 7th International Congress of Phonetic Sciences*; Montreal, Canada. The Hague: Mouton.

Maddieson, I. (1978). Universals of Tone. In J. H. Greenberg, C. Ferguson, & E.A. Moravcsik (Eds.), *Universals of Human Language* (pp. 335-365). Palo Alto: Stanford University Press.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Maddieson, I. (2008). Tone. In M. Haspelmath, M. Dryer, D. Gil, & B. Comrie (Eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library.

Massaro, D.W., Cohen, M.M., & Tseng, C.C. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics*, *13*, 267-289.

Miller-Ockhuizen, A. & Sands, B. (2000). Contrastive Lateral Clicks and Variation in Click Types. In *Proceedings of ICSLP 2000*, *Vol. 2* (pp. 499-500); Beijing, China.

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

Morén, B. & Zsiga, E.C. (2006). The lexical and post-lexical phonology of Thai tones. *Natural Language and Linguistic Theory 24*(1), 113-178.

Newman, P. (1986). Contour tones as phonemic primes in Grebo. In K. Bogers, H. van der Hulst, & M. Mous (Eds.), *The Phonological Representation of Suprasegmentals* (pp. 175-194). Dordrecht: Foris.

Nolan, F. (2007). Intonational equivalence: an experimental evaluation of pitch scales. In *Proceedings of the 15th International Congress of Phonetic Sciences*; Barcelona, Spain, August 3-9, 2003.

Orie, Ọ. (2006). L2 Acquisition and Yoruba Tones: Issues and Challenges. In *Proceedings of the 36th Annual Conference on African Linguistics*; Somerville: Cascadilla Proceedings Project.

Pike, K.L. (1948). Tone Languages; a technique for determining the number and type of pitch contrasts in a language, with studies in tonemic substitution and fusion. Ann Arbor: University of Michigan Press.

Pulleyblank, D. (1986). *Tone in Lexical Phonology*. Dordrecht: Reidel.

R, Free Software Foundation, GNU General Public License, http://www.r-project.org/

Ray, P. S. (1967). Dafla phonology and morphology. *Anthropological Linguistics, 9*(8), 9-14.

Recasens, D. & Espinosa, A. (2006). Dispersion and variability of Catalan vowels. *Speech Communication, 48*(6), 645-666.

Rensch, C. (1968). *Proto-Chinantec phonology*. Mexico City: Museo Nacional de Anthropologia.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime Reference Guide*. Psychology Software Tools, Pittsburgh, PA.

Schroeder, E., Atal, B.S., & Hall, J.L. (1979). Objective measures of certain speech signal degradations based on masking properties of human auditory perception. In B. Lindblom and S. Öhman (Eds.), *Frontiers of Speech Communication Research* (pp. 217-229). London: Academic Press.

Sedlak, P. (1969). Typological considerations of vowels quality systems. *Working Papers on Language Universals, 1* (pp.1-40), Stanford University.

Shen, X. (1990). On Mandarin tone 4. *Australian Journal of Linguistics, 10*, 41-59.

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

Sivertsen, E. (1956). Pitch problems in Kiowa. *International Journal of Applied Linguistics, 22,* 117-130.

Trubetzkoy, N. (1929). Zur allgemeinen Theorie der phonologischen Vokalsysteme. *Travaux* *du Cercle Linguistique de Prague*, *1*.

University of California – Los Angeles Department of Linguistics. (2010). *Praat script for converting Hz to Semitones*. Retrieved July 1, 2010 from http://www.linguistics.ucla.edu/faciliti/facilities/acoustic/pitch_unit_conversion.txt.

Upson, J. (1968). Chatino length and tone. *Anthropological Linguistics, 10,* 1-7.

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

Wan, I.-P. (1999). Mandarin Phonology: Evidence from Speech Errors. PhD dissertation, Department of Linguistics, State University of New York - Buffalo, Buffalo, NY.

Wan, I.-P. (2007). On the phonological organization of Mandarin tones. *Lingua, 117*, 1715-1738.

Wan, I.-P. & J. Jaeger (1998). Speech errors and the representation of tone in Mandarin Chinese. *Phonology, 15*(3), 417-461.

Wang, Y., Jongman, A., & Sereno, J. (2006). L2 acquisition and processing of Mandarin tone. In E. Bates, L. Tan, & O. Tzeng (Eds.), *Handbook of Chinese Psycholinguistics* (pp. 250-257). Cambridge: Cambridge University Press.

Wayland, R. & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, *36*, 250-267.

Weinberger, S. (2010). *Stella elicitation paragraph*. In The Speech Accent Archive, George Mason University. Retrieved from http://accent.gmu.edu/.

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

Wong, P., Schwartz, R.G., & Jenkins, J.J. (2005). Perception and production of lexical tones by 3-year-old, Mandarin-speaking children. *Journal of Speech, Language, and Hearing Research, 48*, 1065-1079.

Wong, P.C.M. & Diehl, R.L. (1998). Effect of spectral center of gravity on vowel height perception. *Journal of the Acoustical Society of America*, *103*, 2981.

Wu, Z.J. (1986). *The Spectrographic Album of Mono-syllables of Standard Chinese.* Beijing: Social Science Press.

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

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica, 55*, 179-203.

Xu, Y. (1999a). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics, 27*, 55-105.

Xu, Y. (1999b). F0 peak delay: When, where and why it occurs. In J. J. Ohala (Ed.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1881-1884); San Francisco, CA.

Yip, M. (1982). Against a segmental analysis of Zahao and Thai: A laryngeal tier proposal. *Linguistic Analysis, 9*, 79-94.

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

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X….& Woodland, P. (2006). *The HTK Book.* Cambridge: Cambridge University Engineering Department.

Yuan, J. & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. In *Proceedings of Acoustics '08* (pp. 5687-5690); Paris, France, June 29-July 4, 2008.

Zee, E. (1978). Duration and intensity as correlates of F0. *Journal of Phonetics*, *6*, 213–220.

Zhang, J. (2002). *The effects of duration and sonority on contour tone distribution: A typological survey and formal analysis*. New York: Routledge.

Zhao, Y. & Jurafsky, D. (2007). The effect of lexical frequency on tone production. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*; August 6-10, 2007, Saarbrücken, Germany.

Zhao, Y. & Jurafsky, D. (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics, 37*, 231–247.

Zsiga, E. & Nitisaroj, R. (2007). Tone features, tone perception, and peak alignment in Thai. *Language and Speech, 50*(30), 343-383.

APPENDIX A: MATERIALS

Cantonese

Cantonese syllables

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
爸	ba	H	*Father*	爸**爸** 爸**爸**去醫院。	*Father* Father goes to hospital.
把	ba	MR	*Hold, Guard*	**把**戲 你唔好 玩**把**戲。	*Trick* Do not play tricks.
霸	ba	M	*Tyranny*	**霸**權 中國外交部 反對 一切**霸**權。	*Hegemony, Tyranny* China's Foreign Ministry is against all forms of hegemony.
---	ba	LR	---	---	---
---	ba	L	---	---	---
罷	ba	F	*Stop*	**罷**工 巴士司機**罷**工維持一個星期。	*Strike* The bus drivers' strike lasts a week.

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
---	bi	H	---	---	---
---	bi	MR	---	---	---
---	bi	M	---	---	---
---	bi	LR	---	---	---
---	bi	L	---	---	---
---	bi	F	---	---	---

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
---	bu	H	---	---	---
---	bu	MR	---	---	---
---	bu	M	---	---	---
---	bu	LR	---	---	---
---	bu	L	---	---	---
---	bu	F	---	---	---

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
打	da	H	*Hit, Beat*	一**打** 唔該一**打**叉燒包。	*Dozen* (quantifier) Please give me a dozen BBQ pork buns.
打	da	MR	*Hit, Beat*	**打**坐 佢每朝都念經**打**坐。	*Meditate* He/she chants and meditates every morning.
---	da	M	---	---	---
---	da	LR	---	---	---
---	da	L	---	---	---
---	da	F	---	---	---

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
---	di	H	---	---	---
---	di	MR	---	---	---
---	di	M	---	---	---
---	di	LR	---	---	---
---	di	L	---	---	---
---	di	F	---	---	---

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
---	du	H	---	---	---
---	du	MR	---	---	---
---	du	M	---	---	---
---	du	LR	---	---	---
---	du	L	---	---	---
---	du	F	---	---	---

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
家	ga	H	*Family, Domestic*	家庭 **家**庭計劃喺戰後推行。	*Family* Family planning was implemented after the war.
假	ga	MR	*Fake; False*	**假** **這隻名錶是假的。**	*Fake, false* This watch brand is fake.
價	ga	M	*Price*	**價**值 呢個意見無**價**值。	*Value* This idea has no value.
---	ga	LR	---	---	---
---	ga	L	---	---	---
嘎	ga	F	*Onomato-poeia*	**嘎**調 ---	*Giggle, Broken sounds (rare)*

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
---	gi	H	---	---	---
---	gi	MR	---	---	---
---	gi	M	---	---	---
---	gi	LR	---	---	---
---	gi	L	---	---	---
---	gi	F	---	---	---

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
姑	gu	H	*Aunt*	**姑**媽 我**姑**媽尋日過身。	*Eldest Aunt* My eldest aunt passed away yesterday.
股	gu	MR	*Thigh, Share*	**股**票 匯控**股**票跌都＄40樓下！	*Stock* Shares of HSBC Holdings are now below $40!
故	gu	M	*Old, Past*	**故**事 小城**故**事多。	*Story* There are lots of stories in a small town.
---	gu	LR	---	---	---
---	gu	L	---	---	---
---	gu	F	---	---	---

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
啦	la	H	*particle of assertion*	**啦啦**隊 我地**啦啦**隊唔差。	*Cheerleading Team* Our cheerleading team isn’t bad.
---	la	MR	---	---	---
罅	la	M	*Crack, Fissure, Split*	**罅**隙 小心**罅**隙月台！	*Gap* Mind the platform gap.
---	la	LR	---	---	---
---	la	L	---	---	---
---	la	F	---	---	---

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
喱	li	H		---	Alternate prounciation of「喱lei」
---	li	MR	---	---	---
---	li	M	---	---	---
---	li	LR	---	---	---
---	li	L	---	---	---
---	li	F	---	---	---

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
---	lu	H	---	---	---
---	lu	MR	---	---	---
---	lu	M	---	---	---
---	lu	LR	---	---	---
---	lu	L	---	---	---
---	lu	F	---	---	---

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
媽	ma	H	*Mother*	媽**媽** 媽**媽**真靚囉！	*Mother* Mom is truly pretty!
媽	ma	MR	*Maid*	**媽** ---	*Maid* (Archaic, Rare)
---	ma	M	---	---	---
麻	ma	LR	*Numb, Hemp*	**麻**醉 今次整容手術用局部**麻**醉。	*Anesthesia* We will use local anesthesia for this plastic surgery.
馬	ma	L	*Horse*	**馬**戲 莫斯科**馬**戲團來港表演。	*Circus* The Moscow Circus comes to Hong Kong.
---	ma	F	---	---	---

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
眯	mi	H	---	**眯** ---	Alternate pronunciation of 眯 [mei]
---	mi	MR	---	---	---
---	mi	M	---	---	---
---	mi	LR	---	---	---
---	mi	L	---	---	---
---	mi	F	---	---	---

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
---	mu	H	---	---	---
---	mu	MR	---	---	---
---	mu	M	---	---	---
---	mu	LR	---	---	---
---	mu	L	---	---	---
---	mu	F	---	---	---

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
哪	na	H	Particle	**哪** ---	Particle (very rare)
---	na	MR	---	---	---
---	na	M	---	---	---
拿	na	LR	*Get, Fetch*	**拿**來 把證件**拿**來。	*Get, Fetch* Get me your ID.
哪	na	L	*What*	**哪**裏 洗手間在**哪**裏?	*Where* Where is the restroom?
那	na	F	Particle	**那** ---	Particle (very rare)

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
呢	ni	H	Particle	呢 ---	Particle (Very rare)
---	ni	MR	---	---	---
---	ni	M	---	---	---
抳	ni	LR	Particle	抳 ---	Particle, used in pointing (very rare)
---	ni	L	---	---	---
---	ni	F	---	---	---

Character	*Syll.*	*Tone*	*Syllable gloss*	*Cantonese noun & phrase*	*Noun/phrase gloss*
---	nu	H	---	---	---
---	nu	MR	---	---	---
---	nu	M	---	---	---
---	nu	LR	---	---	---
---	nu	L	---	---	---
---	nu	F	---	---	---

The North Wind and the Sun translated into Cantonese

有一次，北風同太陽喺度拗緊邊個叻啲。佢哋啱啱睇到有個人行過，嗰個人着住件大褸。佢哋就話嘞，邊個可以整到嗰個人除咗件褸呢，就算邊個叻啲嘞。於是，北風就搏命咁吹。點知，佢越吹得犀利，嗰個人就越係喼實件褸。最後，北風冇晒符，唯有放棄。跟住，太陽出嚟晒咗一陣，嗰個人就即刻除咗件褸嘞。於是，北風唯有認輸啦。

Thai

Thai syllables

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
ปา	สามัญ Neutral	ba	*to throw*	เขา<u>ปา</u>ลูกบอลใส่ตระกร้า	He throws the ball in the basket.
ป่า	เอก Low	bă	*forest*	กรุงเทพไม่มี<u>ป่า</u>	There is no forest in Bangkok.
ป้า	โท Falling	bà	*aunt*	<u>ป้า</u>ของฉันเป็นนักบัญชี	My aunt is an accountant.
ป๊า	ตรี High	bâ	*Chinese way to say father*	<u>ป๊า</u>ชอบอ่านหนังสือพิมพ์ตอนเช้า	My dad likes to read a newspaper in the morning.
ป๋า	จัตวา Rising	bá	*Chinese way to say father*	คนไทยเชื้อสายจีน เรียกพ่อว่า <u>ป๋า</u>	Chinese-Thai people call their dad ป๋า

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
ปี	สามัญ Neutral	bi	*year*	1 <u>ปี</u>มี 12 เดือน	A year has 12 months.
ปี่	เอก Low	bĭ	*flute*	เพื่อนของฉันเป่า<u>ปี่</u>ในวงดนตรีของโรงเรียน	My friend plays flute in the school band.
ปี้	โท Falling	bì	---	---	---
ปี๊	ตรี High	bî	---	---	---
ปี๋	จัตวา Rising	bí	---	---	---

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
ปู	สามัญ Neutral	bu	*crab*	**ปู**นึ่งเป็นอาหารทะเลอร่อยที่สุด	Steamed crab is the best seafood dish.
ปู่	เอก Low	bŭ	*grandfather on father's side*	**ปู่**ของฉันเคยเป็นทหารอากาศ	My grandfather was an Air Force officer.
ปู้	โท Falling	bù	---	---	---
ปู๊	ตรี High	bû	---	---	---
ปู๋	จัตวา Rising	bú	---	---	---

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
ตา	สามัญ Neutral	da	*eye*	ฉันมี<u>**ตา**</u>สีน้ำตาล	I have brown eyes.
ต่า	เอก Low	dă	---	---	---
ต้า	โท Falling	dà	---	---	---
ต๊า	ตรี High	dâ	---	---	---
ต๋า	จัตวา Rising	dá	---	---	---

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
ตี	สามัญ Neutral	di	*to hit*	ครูสมัยก่อนทำโทษนักเรียนโดยการตี	In the past, a teacher punished his students by hitting them.
ตี่	เอก Low	dĭ	*adjective for small eyes*	น้องสาวของฉันตาตี่	My sister has small eyes.
ตี้	โท Falling	dì	---	---	---
ตี๊	ตรี High	dî	---	---	---
ตี๋	จัตวา Rising	dí	*Chinese way to say younger brother*	อาตี๋ไม่สบาย	My younger brother is sick.

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
ตู	สามัญ Neutral	du	---	---	---
ตู่	เอก Low	dŭ	---	---	---
ตู้	โท Falling	dù	*cabinet, cupboard*	แม่ของฉันชอบเก็บของไว้ในตู้	My mom always keeps her stuff in the cabinet.
ตู๊	ตรี High	dû	---	---	---
ตู๋	จัตวา Rising	dú	---	---	---

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
กา	สามัญ Neutral	ga	*a crow,* *a kettle*	ฉันเห็นอีกาเกาะอยู่บนประตู ฉันใช้กาต้มน้ำต้มชา	I saw a crow perching on the gate. I use the kettle to make tea.
ก่า	เอก Low	gă	---	---	---
ก้า	โท Falling	gà	---	---	---
ก๊า	ตรี High	gâ	---	---	---
ก๋า	จัตวา Rising	gá	---	---	---

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
กี	สามัญ Neutral	gi	---	---	---
กี่	เอก Low	gĭ	*how, how many*	คุณเคยมาประเทศไทยกี่ครั้งแล้ว?	How many times have you been to Thailand?
กี้	โท Falling	gì	*hot pot*	ฉันชอบกินสุกี้ เอ็มเค	I love eating MK hot pot.
กี๊	ตรี High	gî	*just now*	เขาเพิ่งมาถึงเมื่อกี๊นี้	He arrived just now.
กี๋	จัตวา Rising	gí	---	---	---

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
กู	สามัญ Neutral	gu	*impolite pronoun to refer to oneself*	กูไม่ชอบกินไอศครีม	I don’t like eating ice cream.
กู่	เอก Low	gŭ	*to holler*	เขากู่ร้องอย่างสุดเสียง	He hollers in his loudest voice.
กู้	โท Falling	gù	*to borrow*	เขากู้เงินจากธนาคาร	He borrows money from a bank.
กู๊	ตรี High	gû	---	---	---
กู๋	จัตวา Rising	gú	---	---	---

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
ลา	สามัญ Neutral	la	*donkey*	ลาเป็นสัตว์ที่หัวรั้นและดื้อดึง	Donkeys have a reputation for stubbornness.
หล่า	เอก Low	lă	---	---	---
ล่า	โท Falling	là	*to hunt*	ลุงของฉันชอบล่าสัตว์	My uncle loves hunting.
ล้า	ตรี High	lâ	*to be tired*	ฉันรู้สึกเหนื่อยล้ามากหลังจากออกกำลังกาย	After working out, I feel very tired.
หลา	จัตวา Rising	lá	*a yard (unit of length)*	1 หลา มี 36 นิ้ว	One yard equals 36 inches.

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
ลี	สามัญ Neutral	li	---	---	---
หลี่	เอก Low	lǐ	---	---	---
ลี่	โท Falling	lì	---	---	---
ลี้	ตรี High	lî	*Li, a Chinese unit of distance*	1 **ลี้** เท่ากับ 300 เมตร	One Li equals about 300 meters.
หลี	จัตวา Rising	lí	---	---	---

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
ลู	สามัญ Neutral	lu	---	---	---
หลู่	เอก Low	lǔ	*to disdain, insult*	คนนิสัยไม่ดีชอบลบ**หลู่**คนอื่น	Only a mean person likes to insult others.
ลู่	โท Falling	lù	*track, path*	โรงเรียนมัธยมของฉันไม่มี**ลู่**วิ่งสำหรับทีมนักวิ่ง	My high school doesn't have a track for the track team.
ลู้	ตรี High	lû	---	---	---
หลู	จัตวา Rising	lú	---	---	---

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
มา	สามัญ Neutral	ma	*to come*	เธอ**มา**จากที่ไหน ?	Where do you come from?
หม่า	เอก Low	mǎ	---	---	---
ม่า	โท Falling	mà	*Chinese way to say grandmother*	อา**ม่า**ของฉันชอบถักไหมพรม	My grandmother loves knitting.
ม้า	ตรี High	mâ	*horse*	ฉันชอบขี่**ม้า**	I love to ride horses.
หมา	จัตวา Rising	má	*dog*	**หมา**ที่บ้านฉันสีน้ำตาล	My dog at home is brown.

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
มี	สามัญ Neutral	mi	*to have*	ฉัน**มี**พี่น้องสองคน	I have two siblings.
หมี่	เอก Low	mĭ	*noodle*	ฉันชอบบะ**หมี่**	My favorite type of noodle is the egg noodle.
มี่	โท Falling	mì	---	---	---
มี้	ตรี High	mî	---	---	---
หมี	จัตวา Rising	mí	bear	ฉันชอบตุ๊กตา**หมี**	I like teddy bears.

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
มู	สามัญ Neutral	mu	---	---	---
หมู่	เอก Low	mŭ	*group, collection*	อาหารมี 5 **หมู่**	There are five different food groups.
มู่	โท Falling	mù	---	---	---
มู้	ตรี High	mû	---	---	---
หมู	จัตวา Rising	mú	*pig, pork*	คนนับถือศาสนาอิสลามไม่รับประทาน**หมู**	Muslims don’t eat pork.

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
นา	สามัญ Neutral	na	*rice field*	ชาวนาปลูกข้าวในทุ่ง**นา**	Farmers grow rice in rice fields.
หน่า	เอก Low	nă	*custard apple*	น้อย**หน่า**เป็นผลไม้เมืองร้อน	Custard apple is one of the tropical fruits.
หน้า	โท Falling	nà	*face*	ผู้หญิงส่วนใหญ่ใส่ใจดูแล**หน้า**มากกว่าผู้ชาย	Most girls take care of their facial skin more than guys do.
น้า	ตรี High	nâ	*aunt*	**น้า**ของฉันอาศัยอยู่ที่ญี่ปุ่น	My aunt lives in Japan.
หนา	จัตวา Rising	ná	*thick*	หนังสือเล่มนี้**หนา**มาก	This textbook is so thick.

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
นี	สามัญ Neutral	ni	---	---	---
หนี่	เอก Low	nǐ	*frugal*	คนตระ**หนี่**ถี่ถ้วนสามารถออมเงินได้มาก	Frugal people can save a lot of money.
หนี้	โท Falling	nì	*debt*	ไม่มีใครอยากยุ่งเกี่ยวกับ**หนี้**สิน	No one wants to deal with debt.
นี้	ตรี High	nî	*this*	กระเป๋าใบ**นี้**ของใคร ?	Whose bag is this?
หนี	จัตวา Rising	ní	*escape, run away from*	ไม่มีใครสามารถ**หนี**จากความจริงได้	No one can escape the truth.

Thai spelling	*Tone* เสียง	*Syll.*	*Syllable gloss*	*Thai phrase*	*Phrase gloss*
นู	สามัญ Neutral	nu	---	---	---
หนู่	เอก Low	nŭ	---	---	---
นู่	โท Falling	nù	---	---	---
นู้	ตรี High	nû	---	---	---
หนู	จัตวา Rising	nú	*mouse*	**หนู**เป็นสัตว์ที่สกปรก	A mouse is a dirty animal.

<u>*The North Wind and the Sun* translated into Thai</u>

ขณะที่ลมเหนือและพระอาทิตย์กำลังเถียงกันว่าใครจะมีพลังมากกว่ากัน ก็มีนักเดินทางผู้หนึ่งเดินผ่านมา ใส่เสื้อกันหนาว ลมเหนือและพระอาทิตย์จึงตกลงกันว่า ใครที่สามารถทำให้นักเดินทางผู้นี้ถอดเสื้อกันหนาวออกได้สำเร็จก่อนจะถือว่าเป็นผู้ที่มีพลังมากกว่า และแล้วลมเหนือก็กระพือพัดอย่างสุดแรง แต่ยิ่งพัดแรงมากขึ้นเพียงใด นักเดินทางก็ยิ่งดึงเสื้อกันหนาวให้กระชับกับตัวมากขึ้นเพียงนั้น และในที่สุดลมเหนือก็เลิกล้มความพยายาม จากนั้นพระอาทิตย์จึงสาดแสงอันร้อนแรงออกมา นักเดินทางก็ถอดเสื้อกันหนาวออกทันที ในที่สุดลมเหนือจึงจำต้องยอมรับว่าพระอาทิตย์มีพลังมากกว่าตน

<u>Mandarin</u>

<u>Mandarin syllables</u>

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
八	bā	ba1	*eight/8*	我有**八**个本子。	I have eight notebooks.
跋	bá	ba2	*to travel; to walk*	经过两天的**跋**涉，队伍终于抵达了目的地。	After traveling two days, the army arrived at the destination.
把	bǎ	ba3	*(measure word)*	门外有两**把**椅子。	There are two chairs outside of the room.
爸	bà	ba4	*father*	**爸爸**回来了。	Father is back.

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
逼	bī	bi1	*to force; to compel*	他们**逼**迫小李离开。	They forced Mr. Li to leave.
鼻	bí	bi2	*nose*	他的**鼻**子在流血。	His nose is bleeding.
笔	bǐ	bi3	*pen; pencil*	桌上有一支钢**笔**。	There is a pen on the table.
币	bì	bi4	*money*	货**币**是一种交换工具。	Money is an exchange tool.

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
---	bū	bu1	---	---	---
醭	bú	bu2	*mold on liquids*	醋上长了白色的**醭**。	There is white mold in the vinegar.
捕	bǔ	bu3	*to catch*	猎人们**捕**获了三只狼。	The hunters caught three wolves.
不	bù	bu4	*no; not*	这样**不**行。	This is not OK.

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
搭	dā	da1	*travel (on boat/train)*	我们**搭**火车去北京。	We traveled to Beijing by train.
达	dá	da2	*to arrive; to achieve*	他们下午三点到**达**北京。	They arrive in Beijing at 3 pm.
打	dǎ	da3	*to fight; to strike*	有人在街上**打**架。	There are people fighting in the street.
大	dà	da4	*big; huge*	房间里有一张**大**桌子。	There is a big table in the room.

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
低	dī	di1	*low; beneath*	这个商场里有很多低价手机。	There are many low-priced cell phones in this department store.
敌	dí	di2	*enemy*	**敌**人被打败了。	The enemies were defeated.
底	dǐ	di3	*bottom*	碗**底**有一粒米。	There is rice on the bottom of the bowl.
地	dì	di4	*earth; ground*	书掉在**地**上了。	The book was dropped on the ground.

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
都	dū	du1	*major city*	中国的首**都**是北京。	Beijing is the capital city of China.
椟	dú	du2	*cabinet; case*	老师给我们说了买**椟**还珠的故事。	The teacher told us a story in which someone bought a diamond's case but returned the diamond.
睹	dǔ	du3	*to observe; to see*	人们都目**睹**了这一历史时刻。	Everyone witnessed this historic moment.
肚	dù	du4	*belly*	他在**肚**子上刺青。	He has a tattoo on his belly.

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
---	gā	ga1	---	---	---
嘎	gá	ga2	*crackling sound; quack*	池塘里的鸭子**嘎嘎**叫。	The ducks in the pond are quacking.
---	gǎ	ga3	---	---	---
尬	gà	ga4	*embarrassed*	他看起来很尴**尬**。	He seemed very embarrassed.

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
---	gī	gi1	---	---	---
---	gí	gi2	---	---	---
---	gǐ	gi3	---	---	---
---	gì	gi4	---	---	---

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
估	gū	gu1	*estimate*	他**估**计股市会大跌。	Based on his estimates, the stock market is going to slump.
---	gú	gu2	---	---	---
古	gǔ	gu3	*ancient; old*	他喜欢读**古**文。	He likes reading books of classical Chinese.
固	gù	gu4	*hard; strong; firm*	这个房子很坚**固**。	This is a strong building.

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
拉	lā	la1	*to pull or drag*	他在**拉**车。	He is pulling the cart.
---	lá	la2	---	---	---
喇	lǎ	la3	*(phonetic)*	**喇**嘛辩论是传统。	Lama debating has a long tradition.
蜡	là	la4	*candle; wax*	库房里有很多**蜡**烛。	There are many candles in the warehouse.

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
喱	lī	li1	*curry*	他喜欢吃咖喱。	He likes curry.
梨	lí	li2	*pear*	他喜欢吃**梨**子。	He likes pears.
娌	lǐ	li3	*husband's brother's wife*	他们妯娌感情很好。	The sisters-in-law have a good relationship.
历	lì	li4	*calendar*	桌上有一本日**历**。	There is a calendar on the table.

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
噜	lū	lu1	*to snore*	他睡觉打呼**噜**。	He snores.
炉	lú	lu2	*stove*	房间里有**炉**子。	There is a stove in the room.
卤	lŭ	lu3	*marinate*	**卤**肉很好吃。	Marinated pork is very tasty.
录	lù	lu4	*to record*	他喜欢把讲座**录**下来。	He likes recording the lectures.

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
妈	mā	ma1	*mother*	她很想**妈妈**。	She misses her mom.
麻	má	ma2	*hemp*	桌上有一条**麻**绳。	There is a hemp rope on the table.
马	mă	ma3	*horse*	他喜欢看赛**马**。	He likes horse racing.
骂	mà	ma4	*scold*	**骂**人解决不了问题。	Scolding people solves nothing.

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
咪	mī	mi1	*(sound to call a cat)*	他有一只猫咪。	He has a little cat.
迷	mí	mi2	*confused*	他看起来很**迷**惑的样子。	He looks very confused.
米	mĭ	mi3	*rice*	我喜欢吃**米**饭。	I like rice.
密	mì	mi4	*secret*	这是一个秘**密**。	This is a secret.

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
---	mū	mu1	---	---	---
---	mú	mu2	---	---	---
母	mŭ	mu3	*mom*	她的**母**亲很漂亮。	Her mom is very beautiful.
募	mù	mu4	*to recruit; to raise*	他们**募**集了100万美元。	They raised a million dollars.

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
---	nā	na1	---	---	---
拿	ná	na2	*to hold; to take*	他手里**拿**着一本书。	He is holding a book in his hands.
哪	nă	na3	*how; which*	我不知道**哪**个菜好吃。	I don't know which dish is tasty.
那	nà	na4	*that; those*	他选了**那**本书。	He picked that book.

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
妮	nī	ni1	*girl*	小**妮**子很可爱。	The little girl is very cute.
泥	ní	ni2	*mud*	他鞋上有**泥**。	There is mud on his shoes.
你	nǐ	ni3	*you*	**你**昨天睡的好么？	Did you sleep well last night?
匿	nì	ni4	*hide*	犯人**匿**藏在山洞里。	Criminals are hiding in the cave.

Character	*PinYin*	*Tone #*	*Syllable gloss*	*Mandarin phrase*	*Phrase gloss*
---	nū	nu1	---	---	---
奴	nú	nu2	*slave*	**奴**隶的生活很凄惨。	Slaves have miserable lives.
努	nǔ	nu3	*to exert; to strive*	他学习很 **努**力。	He studies hard.
怒	nù	nu4	*indignant; furious*	他们都很愤**怒**。	They are all furious.

The North Wind and the Sun translated into Mandarin

有一次，**北**风和太阳在争论谁更强大，这时一个穿着大衣的行人正巧经过。太阳和北风决定，谁让这个行人脱掉大衣谁就更强大。于是，北风拼命吹风，但是风越大，这个行人把大衣裹得越紧，最终北风只好放弃。接下来，太阳放射出温暖的阳光，这个行人马上就脱掉了大衣。于是北风不得不服输，承认太阳更强大。

Yoruba

Yoruba syllables

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
bá	H	*To meet*	Mo fe lo ba Tunde.	I am going to meet Tunde.
ba	M	---	---	---
bà	L	*To hit*	So oko ba eiye.	Throw a stone at the bird.

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
bí	H	*To ask*	Kini o bi mi fun?	Why are you asking me?
bi	M	---	---	---
bì	L	*To vomit*	Kini o de to nfi bi?	Why are you vomiting?

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
bú	H	*To fetch*	Lo bu omi wa.	Go and fetch water.
bu	M	*To curse*	Ye bu mi mo.	Do not curse me again.
bù	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
dá	H	*To break*	Ma da igi yen.	Do not break that stick.
da	M	---	---	---
dà	L	*To spill*	Mo ti da omi nu.	I spilled the water.

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
dí	H	*To weave*	Mo nlo di irun mi.	I am going to weave my hair.
di	M	---	---	---
dì	L	*To hold*	Ma di mi mu.	Do not hold me.

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
dú	H	---	---	---
du	M	---	---	---
dù	L	*To rush*	Won du oko wo.	They rushed into the bus.

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
gá	H	*Height*	Ta lo ga ju?	Who is the tallest?
ga	M	---	---	---
gà	L	*To choke*	O ga mi lorun.	He choked me.

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
gí	H	---	---	---
gi	M	---	---	---
gì	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
gú	H	---	---	---
gu	M	---	---	---
gù	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
lá	H	*To lick*	Mo fe la oyin.	I want to lick some honey.
la	M	---	---	---
là	L	*To dream*	Mo la ala kan.	I had a dream.

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
lí	H	---	---	---
li	M	---	---	---
lì	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
lú	H	*To beat*	Ye lu mi.	Stop beating me.
lu	M	---	---	---
lù	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
má	H	---	---	---
ma	M	(pronoun)	Ma je ounje yen.	Do not eat that food.
mà	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
mí	H	(pronoun)	Mi o binu.	I am not angry.
mi	M	---	---	---
mì	L	*To swallow*	Gbe ogun yen mi.	Swallow the medicine.

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
mú	H	*To bring*	Lo mu owo wa.	Go and bring money.
mu	M	*To drink*	Ye mu oti mo.	Do not drink beer again.
mù	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
ná	H	*To spend*	Ni na ni owo	Spending money
na	M	---	---	---
nà	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
ní	H	*To own*	Ta lo ni moto?	Who owns this car?
ni	M	---	---	---
nì	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Yoruba phrase*	*Phrase gloss*
nú	H	---	---	---
nu	M	*To hand-feed*	Mo fe nu omo mi.	I want to hand-feed my baby.
nù	L	---	---	---

The North Wind and the Sun translated into Yoruba

Ni ojo kan Afefe ati Orun nleri eniti o lagbara ju,won ri arinrin ajo kan ti o wo ewu otutu.Won wa pinu pe eniti o ba koko mu arinrin ajo na bo aso otutu ti wo ni o ni agbara ju. Ni oju ese, Afefe ba bere si ni fe.Afefe na ni agbara gan ni,sugbon kaka ki arinrin ajo bo aso otutu,nise ni otun wa mo ara re.Ni igba to ya, o re Afefe o ba ni ohun jawo.Lehin na Orun ba bere si ni ran, ni ojukana ni arinrin ajo ba bo aso otutu ti o wo.Afefe ba jewo pe Orun ni oni agbara ju ninu awon mejeji.

Igbo

Igbo syllables

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
bá	H	*auxiliary to be*	I **bá** úbá jí átú ùtó.	To be rich gives joy.
bà	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
bí	H	*to live*	Ó **bí** n'úló áhù.	He lives in that house.
bì	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
bú	H	*be/are/is*	Í **bú** nwókè.	You are a man.
bù	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
dá	H	*to warm*	Ó gá **dá** nrí áhù n'ókú.	He will warm up the food.
dà	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
dí	H	*husband*	Ó bú **dí** m.	He is my husband.
dì	L	*to exist*	Ágbàlí úkù áhù **dì** n’ebe à.	That shoe is here.

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
dú	H	*a variation in dialect that means to establish*	Nwókè áhù **dú**zírì nwá ahu.	That man established that child beautifully.
dù	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
gá	H	---	---	---
gà	L	*auxiliary indicating future action*	Ó **gà** á lótà úló.	He will come home.

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
gí	H	*you*	Ó bú **gí** nyérèm áká.	It was you who helped me.
gì	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
gú	H	---	---	---
gù	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
lá	H	---	---	---
là	L	*to go, to leave*	Á ná m à **là** ngàm.	I am going to my place.

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
lí	H	---	---	---
lì	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
lú	H	---	---	---
lù	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
má	H	---	---	---
mà	L	*nothing,* *either/or,* *if,* *but*	-Ó gàghí áfó **mà** ótù. -**Mà** òbù gí **mà** òbù yá. -Ó gà émè yá **mà** ó nyè yá égò. -Ó gà ékwé **mà** ó gà ététù áka.	-Nothing will be left. -Either you or him. -He will do it only if paid. -He will agree but it will take a while.

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
mí	H	---	---	---
mì	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
mú	H	*me*	Ó bú **mú**.	It is me.
mù	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
ná	H	*without*	Ó méré yá **<u>n'á</u>**mághí ámà.	He did it without knowing.
nà	L	*and,* *that*	-Ádá **<u>nà</u>** Óbí. -Ó bù íhé **<u>nà</u>** éméme.	-Ada and Obi. -That is true.

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
ní	H	---	---	---
nì	L	---	---	---

Syllable	*Tone*	*Syllable gloss*	*Igbo phrase*	*Phrase gloss*
nú	H	---	---	---
nù	L	---	---	---

<u>*The North Wind and the Sun* translated into Igbo</u>

Ìkùkù ṹgṹrụ̀ nà Ánwụ̄ nà-arṹrị̄tá ṹkà ónyé kā ībè yá íké m̀gbè há hụ̀rụ̀ ótù ónyé íjè kà ó yì ùwé ṹgụ̀rụ̀ yá nà-àbịá. Há kwèkọ̀rị̀tàrà nà ónyé būrū ṹzọ̀ méé kà ónyé íjè áhụ̀ yípụ̀ ùwé yā kà á gà-éwè dị́kà ónyé ka íbè yá íké. Ìkùkù ṹgụ̀rụ̀ wéé mà̀lị́té féé, féé, òtù íké yā hà; mà kà ọ́ nà-èfé kà ónyé íjè áhụ̀ nà-èjị́désí ūwē yā īkē nà àhú yā. Yá fékàtá hápụ̀. M̀gbè áhụ̀ Ánwụ̄ wéé chápụ̀tá, chásíkē, méé kà ébé níílē kpòró ọ́kụ̄; ná-ātùfūghì ógè ónyé íjè áhụ̀ yìpụ̀rụ̀ ùwé yā. Ǹké à mèrè ìkùkù ṹgụ̀rụ̀ kwèrè nà Ánwụ̄ kà yá íké.

<u>*Stella* passage</u>

Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.

APPENDIX B: PARTICIPANTS

ID	Age (ys)	Lg	Block Order		1st town, age there		2nd town, age there		3rd town, age there		4th town, age there	
			1	2								
CF02	22	C	R	S	HK	0-19	Ev	19-22				
CF03	18	C	S	R	HK	0-18	Ev	18-				
CF04	30	C	S	R	Guangdong province	0-5	HK	5-29	Ev	29-		
CM02	29	C	R	S	Pittsburgh, PA	0-2	HK	2-27	Ev	27-		
CM03	30	C	S	R	HK	0-30	Ev	30-				
CM04	22	C	R	S	HK	0-22	Ev	22-				
TF01	20	T	R	S	Bk	0-18	Ev	18-				
TF04	26	T	S	R	Bk	1-17, 18-23	Farmville, VA	17-18				
TF05	25	T	S	R	Bk	0-25	Chicago	25-				
TM02	27	T	R	S	Bk	0-26	Ev	26-27				
TM04	19	T	S	R	Bk	0-15	Lawrenceville, NJ	15-18				
TM05	31	T	S	R	Bk	0-17	Newport, RI	17-18				
MF02	20	Mn	R	S	Bei	0-18	Ev	18-20				
MF03	23	Mn	S	R	Bei	0-23	Ev	23-				
MF05	19	Mn	S	R	Bei	0-13	Saratoga, CA	13-15	Shanghai, China	15-19	Ev	19-
MM02	24	Mn	R	S	Bei	0-19	HK	19-20, 21-22	San Diego, CA	20-21		
MM03	28	Mn	S	R	Bei	0-19	Nanjing, China	19-23	Ev	23-28		
MM04	26	Mn	S	R	Bei	0-23	Ev	23-26				
YF03	45	Y	R	S	U	U						
YF05	47	Y	S	R	Abeokuta, N	0-7	Lagos or Ibadan, N	7-32	Chicago	32-		
YF07	28	Y	R	S	(U), N	0-18	Chicago	18-				
YM02	34	Y	R	S	Ibadan, N	0-24	Lagos or Port-Harcourt, N	24-27	Stillwater, OK	27-29		
YM05	46	Y	S	R	Lagos, N	0-19	Chicago	19-				
YM06	42	Y	S	R	U	U						
IF02	39	I	R	S	Lagos, N	1-10, 18-21	Owerri, N	10-18	London, England	21-23	Chicago	23-
IF04	50	I	R	S	Lagos, N	0-33	Chicago	33-				
IF05	28	I	S	R	U	U						
IM04	45	I	R	S	Aba, N	0-18, 21-44	Lagos, N	18-21	Chicago	44-		
IM05	33	I	R	S	Nsukka, N	0-28	Chicago	28-				
IM07	42	I	S	R	Enugu, N	0-21	Lagos, N	22-28	Chicago	29-		

Table B1. Participant demographic information

Legend										
Language			**Sex**			**Block order**			**Town/country**	
Code	***Gloss***		***Code***	***Gloss***		***Code***	***Gloss***		***Code***	***Gloss***
Lg	Language		F	Female		S	Sequential		U	Unreported or unknown
C	Cantonese		M	Male		R	Random		Ev	Evanston, IL, USA
T	Thai								HK	Hong Kong, China
Mn	Mandarin								Bk	Bangkok, Thailand
Y	Yoruba								Bei	Beijing, China
I	Igbo								N	Nigeria

Table B2. Legend for Table B1

Cantonese						
Talker	**CF02**	**CF03**	**CF04**	**CM02**	**CM03**	**CM04**
Mean F0 (ST)	10.561	10.326	11.044	3.285	6.824	3.059
Median F0 (ST)	10.403	10.998	10.884	2.541	5.913	2.561
Min F0 (ST)	-2.474	-4.918	-4.663	-4.982	-2.275	-4.929
Max F0 (ST)	15.823	27.318	22.765	24.224	27.901	30.416
Range F0 (ST)	18.297	32.236	27.427	29.206	30.176	35.344
Thai						
Talker	**TF01**	**TF04**	**TF05**	**TM02**	**TM04**	**TM05**
Mean F0 (ST)	10.839	9.809	10.922	5.252	4.488	2.8
Median F0 (ST)	11.179	11.112	11.042	4.977	4.117	1.635
Min F0 (ST)	-4.934	-4.9544	-4.763	0.513	-4.9847	-4.433
Max F0 (ST)	25.892	24.883	25.719	16.102	27.205	25.625
Range F0 (ST)	30.826	29.837	30.482	15.589	32.189	30.058
Mandarin						
Talker	**MF02**	**MF03**	**MF05**	**MM02**	**MM03**	**MM04**
Mean F0 (ST)	12.252	13.502	12.011	7.674	-0.1369	2.27
Median F0 (ST)	13.581	14.536	12.884	8.202	-0.011	2.491
Min F0 (ST)	-4.862	-4.8648	-4.857	-4.921	-4.9874	-4.944
Max F0 (ST)	28.526	29.356	23.888	26.814	7	11.378
Range F0 (ST)	33.388	34.221	28.745	31.735	11.987	16.323
Yoruba						
Talker	**YF03**	**YF05**	**YF07**	**YM02**	**YM05**	**YM06**
Mean F0 (ST)	11.568	9.69	11.048	3.539	-0.52	5.828
Median F0 (ST)	11.536	9.49	11.422	4.125	-0.723	5.922
Min F0 (ST)	1.44	-1.9868	5.676	-4.8	-4.958	-4.81
Max F0 (ST)	22.169	21.294	18.413	8.848	10.059	12.922
Range F0 (ST)	20.729	23.281	12.737	13.648	15.017	17.732
Igbo						
Talker	**IF02**	**IF04**	**IF05**	**IM04**	**IM05**	**IM07**
Mean F0 (ST)	11.948	11.156	14.489	6.749	3.572	2.155
Median F0 (ST)	11.744	11.528	14.41	6.094	4.319	3.107
Min F0 (ST)	-0.604	2.87	5.319	-3.637	-4.5731	-4.858
Max F0 (ST)	20.125	18.217	22.149	23.916	8.751	8.428
Range F0 (ST)	20.729	15.347	16.83	27.553	13.324	13.286

Table B3. Descriptive statistics of participants' tone productions

APPENDIX C: INSTRUCTIONS (in Cantonese)

Hello!
你好!

Please press the space bar to begin.
請按空格鍵開始。

Thank you for participating in this study!
多謝你參與呢個研究。

Please read the following instructions carefully.
請小心閱讀以下指示。

If you have any questions, at any time, please ask the experimenter.
如果你喺任何時間有任何問題，請詢問研究員。

Today you will be recorded as you read some syllables aloud.
今日實驗會錄低你朗讀嘅音節。

Language researchers understand that in Cantonese, the pitch of the word indicates the meaning of the word, and that each of these different pitch types is called a tone.
語言研究指出廣東話中嘅音高（稱為音調）顯示每個字嘅意思。

We know that Cantonese has six different tones.
廣東話有六個音調。

We know that a syllable – for instance, *si* – can be produced with each of these six tones, and that each means something different.
一個音節(例如*si*) 可以用六個音調讀出，而意思不同。

To continue with the example *si*, we know that:
以*si*舉例,

- The word [詩] (the syllable *si* pronounced with the high-level tone) means poetry.

*si*以陰平調讀出，喺 詩。

- The word [史] (the syllable *si* pronounced with the mid-rising tone) means history.

*si*以陰上調讀出，喺 史。

- The word [試] (the syllable *si* pronounced with the mid-level tone) means to try.

si 以陰去調讀出，喺 試。

- The word [時] (the syllable *si* pronounced with the low-rising tone) means time.

si 以陽平調讀出，喺 時。

- The word [市] (the syllable *si* pronounced with the low-level tone) means market.

si 以陽上調讀出，喺 市。

- The word [視] (the syllable *si* pronounced with the mid-falling tone) means to watch.

si 以陽去調讀出，喺 視。

During the experiment, you will see charts that each describe a single syllable, e.g.,

Chinese Character	*Tone #*	*Tone description*	*Meaning of character (in English)*	*Noun; phrase (in Cantonese)*	*Meaning of noun and phrase (in English)*
詩	si1	High-Level	*poetry*	佢熟讀唐**詩**三百首。	He pretty much memorized the 300 poems from the Tang dynasty.

On the far left, under the Chinese character heading, is this word written in Chinese.
以中文書寫嘅字喺左面 Chinese character 以下。

Next, under the Tone # heading, is the syllable written with its tone number.
然後，音調編號喺Tone # 以下。

Next, under the Tone description heading, is a description of the tone.
然後，音調喺Tone description 以下。

Next, under the Meaning of character (in English) heading, is the meaning of the character in English.
然後，單字嘅英文意思喺 Meaning of character (in English) 以下。

Next, under the Noun; phrase (in Cantonese) heading, is the character as it occurs in a Cantonese noun, and that noun used in a short Cantonese example phrase.
然後，單字構成嘅廣東話詞語, 同埋呢個詞語構成嘅廣東話例句喺 Noun; phrase (in Cantonese) 以下。

Finally, under the Meaning of noun and phrase (in English) heading, is the translation of the Cantonese noun and phrase.
最後，廣東話詞語同例句嘅英文翻譯喺 Meaning of noun and phrase (in English) 以下。

You will see one of these charts per screen.
你會喺每一個螢幕見到其中一個表。

Your job: Please read the syllable aloud. Speak as clearly as possible, and only say it ONCE.
你要清楚朗讀呢一個音節一次。

For this example, you would simply say [詩] (the syllable *si* pronounced with the high-level tone).
例如，你只需讀出[詩](*si*以陰平調讀出)。

You can have as much time as you need to think about the word before you say it.
讀音之前，你會有時間諗清楚。

When you're finished, hit the space bar to continue.
結束後，請按空格鍵繼續。

Got it? Let's try a few examples.
清楚未？請嘗試幾個例子。

When you're ready for the first example, hit the space bar.
當你準備好，請按空格鍵繼續。

Chinese Character	*Tone #*	*Tone description*	*Meaning of character (in English)*	*Noun; phrase (in Cantonese)*	*Meaning of noun and phrase (in English)*
史	si2	Mid-Rising	*history*	《**史**記》是司馬遷編寫的一本歷**史**著作。	*Shiji* is a historical work written by Sima Qian.

Please read this syllable aloud. Speak clearly, and only say it once. Then press the space bar to continue.
請清楚發出呢一個音節一次，然後按空格鍵繼續。

Chinese Character	*Tone #*	*Tone description*	*Meaning of character (in English)*	*Noun; phrase (in Cantonese)*	*Meaning of noun and phrase (in English)*
試	si3	Mid-Level	*to try*	政府嘗**試**禁酒。	The government attempts to ban alcohol.

Please read this syllable aloud. Speak clearly, and only say it once. Then press the space bar to continue.
請清楚發出呢一個音節一次，然後按空格鍵繼續。

Chinese Character	*Tone #*	*Tone description*	*Meaning of character (in English)*	*Noun; phrase (in Cantonese)*	*Meaning of noun and phrase (in English)*
時	si4	Low-Rising	*time*	**時**間寶貴。	Time is precious.

Please read this syllable aloud. Speak clearly, and only say it once. Then press the space bar to continue.
請清楚發出呢一個音節一次，然後按空格鍵繼續。

Chinese Character	*Tone #*	*Tone description*	*Meaning of character (in English)*	*Noun; phrase (in Cantonese)*	*Meaning of noun and phrase (in English)*
市	si5	Low-Level	*market*	我主修**市**場學。	I major in marketing.

Please read this syllable aloud. Speak clearly, and only say it once. Then press the space bar to continue.
請清楚發出呢一個音節一次，然後按空格鍵繼續。

Chinese Character	*Tone #*	*Tone description*	*Meaning of character (in English)*	*Noun; phrase (in Cantonese)*	*Meaning of noun and phrase (in English)*
視	si6	Mid-Falling	*To watch*	電**視**正在進行著一場革命。	Television is undergoing a revolution.

Great job! Press space to continue.
好，請按空格鍵繼續。

The examples you just saw involved real, meaningful, words.
以上都係有意思嘅字。

Sometimes, you will be asked to produce meaningless syllables (non-words).
有時候，我要你發音，但係你嘅發音係無意思。

In the chart for non-words, you'll see a series of dashes (----) under most of the headings.
係表裡面，你會見到 ----。

For instance, the syllable *ki* produced with the high-level tone is not a meaningful word in Cantonese:

例如，以陰平調發出嘅音節ki唔係一個廣東話字。

Chinese Character	*Tone #*	*Tone description*	*Meaning (in English)*	*Example sentence (in Chinese)*	*English translation*
----	ki1	High-Level	----	----	----

Here, your job is the same: Please read the syllable aloud. Speak as clearly as possible, and only say it ONCE.
同樣，你要清楚發出呢一個音節一次。

For this example, you would say the syllable *ki* spoken with the high-level tone.
以呢個例子，你都係以陰平調發出音節ki。

Again, you can have as much time as you need to think about the syllable before you say it.
同樣，讀音之前，你會有時間諗清楚。

When you're finished saying the syllable, hit the space bar to continue.
結束後，請按空格鍵繼續。

Got it? Let's try a few examples.
清楚未？請嘗試幾個例子。

When you're ready for the first example, hit the space bar.
當你準備好，請按空格鍵繼續。

Chinese Character	*Tone #*	*Tone description*	*Meaning (in English)*	*Example sentence (in Chinese)*	*English translation*
----	ki2	Mid-Rising	----	----	----

Please read this syllable aloud. Speak clearly, and only say it once. Then press the space bar to continue.
請清楚發出呢一個音節一次，然後按空格鍵繼續。

Chinese Character	*Tone #*	*Tone description*	*Meaning (in English)*	*Example sentence (in Chinese)*	*English translation*
----	ki3	Mid-Level	----	----	----

Please read this syllable aloud. Speak clearly, and only say it once. Then press the space bar to continue.
請清楚發出呢一個音節一次，然後按空格鍵繼續。

Chinese Character	*Tone #*	*Tone description*	*Meaning (in English)*	*Example sentence (in Chinese)*	*English translation*
----	ki4	Low-Rising	*market*	----	----

Please read this syllable aloud. Speak clearly, and only say it once. Then press the space bar to continue.
請清楚發出呢一個音節一次，然後按空格鍵繼續。

Chinese Character	*Tone #*	*Tone description*	*Meaning (in English)*	*Example sentence (in Chinese)*	*English translation*
----	ki5	Low-Rising	----	----	----

Please read this syllable aloud. Speak clearly, and only say it once. Then press the space bar to continue.
請清楚發出呢一個音節一次，然後按空格鍵繼續。

Chinese Character	*Tone #*	*Tone description*	*Meaning (in English)*	*Example sentence (in Chinese)*	*English translation*
----	ki6	Low-Rising	----	----	----

Great job! Press space to continue.
好，請按空格鍵繼續。

So, your job is simply to read each syllable aloud, just once, as clearly as possible.
你只需要清楚朗讀每一個音節一次。

After you read a number of them, the computer program will stop.
每當讀完一連串音節，電腦程式會停頓。

You will read lists of these syllables six times total. Be aware that, in three of these lists, meaningful syllables and meaningless syllables will appear in random order.
你會讀出一連串音節合共六次。喺其中三個系列，電腦會顯示隨機有意思同無意思嘅音節。

After you read the syllables, you will be asked to read two short passages: one in Cantonese and one in English.
朗讀完畢後，你會用廣東話同英文讀兩篇短文。

You will be given several breaks throughout this recording session.
錄音時，你會暫停幾次。

Please take a break, and let the experimenter know that the program has stopped.
暫停時，請告知實驗人員電腦程式停頓。

Please read the following passage aloud.
請清楚朗讀以下文章。

When you’re finished, please let the experimenter know.
完畢後，請告知實驗人員。

You’re all finished!
搞掂晒！

APPENDIX D: CALCULATIONS OF CANTONESE, MANDARIN, AND YORUBA TONE-SPACE SIZES AT ONSET, MIDPOINT, AND OFFGLIDE FOR SECTION 4.4

Tables D1, D2, and D3 summarize the *ToneSpace* analysis results at onset, midpoint, and offglide, respectively. In each table, the cell under the "Est" column and in the "Tone B" row refers to the estimated difference between the top and bottom tones for Cantonese. The size of the tone space is the absolute value of this number, rounded to the nearest whole semitone. The value in the cell under the "Est" column and in the "LanguageM:ToneB" refers to the adjustment to the Cantonese estimate that is required to estimate the Mandarin tone space. Likewise, the value in the cell under the "Est" column and in the "LanguageY:ToneB" refers to the adjustment to the Cantonese estimate that is required to estimate the Yoruba tone space.

ToneSpaceOnset: **Cantonese vs. Mandarin and Yoruba**				
	Est	**St.E**	**t-val**	**pMCMC**
ToneB	**-5.53**	0.123	-44.780	0.0001
LanguageM:ToneB	**-2.95**	0.149	-19.870	0.0001
LanguageY:ToneB	**2.07**	0.149	13.920	0.0001

Table D1. *ToneSpaceOnset* results for Cantonese vs. Mandarin and Yoruba

For the purposes of approximating the languages' tone-space sizes for section 4.4, examinations of cross-language tone dispersion, the tone-space size values are rounded to the nearest whole semitone, after adjustments to the Cantonese estimate. As intimated above, the data in Table D1 indicates that the Cantonese tone space at onset is approximately 6 ST wide (the absolute value of the number in the cell under the "Est" column and in the "ToneB" row, rounded to the nearest whole semitone). The Mandarin tone space at onset is approximately 8 ST wide (-5.53 + (-2.95) = -8.48, rounded down to -8 ST, the absolute value of which is 8 ST). The Yoruba tone space at onset is approximately 3 ST wide (-5.53 + 2.07 = -3.46, rounded down to -3 ST, the absolute value of which is 3 ST). It is also important to note that the full set of *ToneSpaceOnset* results showed that the Cantonese, Mandarin, and Yoruba tone spaces are significantly different in size at onset.

ToneSpaceMidpoint: **Cantonese vs. Mandarin and Yoruba**				
	Est	**St.E**	**t-val**	**pMCMC**
ToneB	**-7.82**	0.178	-43.970	0.0001
LanguageM:ToneB	**-1.33**	0.221	-5.990	0.0001
LanguageY:ToneB	**1.95**	0.223	8.750	0.0001

Table D2. ***ToneSpaceMidpoint*** **results for Cantonese vs. Mandarin and Yoruba**

The data in Table D2 indicates that the Cantonese tone space at midpoint is approximately 8 ST wide (the absolute value of the number in the cell under the “Est” column and in the “ToneB” row, rounded to the nearest whole semitone). The Mandarin tone space at onset is approximately 9 ST wide (-7.82 + (-1.33) = -9.15, rounded down to -9 ST, the absolute value of which is 9 ST). The Yoruba tone space at onset is approximately 6 ST wide (-5.53 + 1.95 = -5.87, rounded up to -6 ST, the absolute value of which is 3 ST). As before, it is also important to note that the full set of *ToneSpaceMidpoint* results showed that the Cantonese, Mandarin, and Yoruba tone spaces are indeed significantly different in size at midpoint.

ToneSpaceOffglide: **Cantonese vs. Mandarin and Yoruba**				
	Est	**St.E**	**t-val**	**pMCMC**
ToneB	**-6.94**	0.276	-25.186	0.0001
LanguageM:ToneB	**0.58**	0.387	1.498	0.1366
LanguageY:ToneB	**-0.08**	0.390	-0.207	0.8402

Table D3. ***ToneSpaceOffglide*** **results for Cantonese vs. Mandarin and Yoruba**

The data in Table D3 indicates that the Cantonese tone space at offglide is about 7 ST wide (the absolute value of the number in the cell under the “Est” column and in the “ToneB” row, rounded to the nearest whole semitone). The Mandarin tone space at onset is approximately 6 ST wide (-6.94 + (0.58) = -6.36, rounded down to -6 ST, the absolute value of which is 6 ST). The Yoruba tone space at onset is approximately 7 ST wide (-6.94 + (-0.08) = -7.02), rounded down to -7 ST, the absolute value of which is 7 ST). Thus, the Cantonese and Yoruba tone spaces are both about 7 ST wide, while the Mandarin tone space is approximately 6 ST wide. However, the full set of *ToneSpaceOffglide* results showed that the Cantonese, Mandarin, and Yoruba tone spaces are not significantly different in size at offglide. I therefore assume for the purposes of estimating tone space sizes in section 4.4 that the languages’ tone space sizes all equal the average size of the three spaces: 7 ST wide ((7 + 6 + 7)/3 = 6.66 ST, rounded up to 7 ST).

VITA

Jennifer Alexandra Alexander

Department of Linguistics, Northwestern University
2016 Sheridan Road, Evanston, IL 60208 USA
Email: jenalex@u.northwestern.edu; Cell: 773-259-8803; Office: 847-491-5832
http://gradstudents.wcas.northwestern.edu/~jaa930

Research Interests

Topics:

- Acoustic, perceptual, and articulatory properties of cross-language lexical tone systems
- Perception and production of pitch in music and speech
- Native- and non-native speech-in-noise perception

Methods: Human behavioral research

Dissertation: *The Theory of Adaptive Dispersion and Acoustic-phonetic Properties of Cross-language Lexical-tone Systems*. Committee: Ann Bradlow, Patrick Wong, & Matthew Goldrick.

Education

2003-present **Ph.D.** candidate in Linguistics, Northwestern University, Evanston, IL

- 2003-present: Cognitive Science specialist
- 2004-present: Language, Music, and Communication specialist
- Date of completion: December 2010

Su04 NU New TA Training Program
Su04 NU ESL Program ESL Tutor and Instructor Training Program
2002 **B.A.** in Linguistics & **B.A.** in English, University of California, Berkeley

Professional Skills

Programming: R, PRAAT, E-Prime
Natural languages: English (native), French (intermediate), Mandarin Chinese (elementary)

Peer-reviewed Publications & Manuscripts in Preparation

Alexander, J.A., Bradlow, A.R., Ashley, R.D., and Wong, P.C.M. (In preparation) On the interaction of speech-pitch processing and music-pitch processing.

Bradlow, A.R. and **Alexander, J.A**. (2007) Semantic-contextual and acoustic-phonetic cues for English sentence-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, 121(4), 2339-2349.

Alexander, J.A., Wong, P.C.M., and Bradlow, A. (2005) Lexical Tone Perception in Musicians and Non-musicians. Proceedings of Interspeech 2005 – Eurospeech – 9th European Conference on Speech Communication and Technology.

Invited Talks

Alexander, J.A., Wong, P.C.M., Bradlow, A.R., and Ashley, R.D. Music Contour Perception in Tone-language- and Non-tone-language Speakers. Northwestern University Cognitive Science Program Cognitive Science Fest 2006. Evanston, IL, May 30, 2006.

Alexander, J.A., Wong, P.C.M., Bradlow, A.R. Lexical Tone Perception and Production in Musicians and Non-musicians. Northwestern University Music Cognition Program colloquium series. Evanston, IL, Nov. 16, 2005.

Refereed Conference Presentations (not in proceedings)

Alexander, J.A., Bradlow, A.R., Ashley, R.D., and Wong, P.C.M. Music melody perception in tone-language and non-tone-language speakers. 156th Meeting of the Acoustical Society of America, Miami, FL, Nov. 10-14, 2008.

Alexander, J.A. and Clark, B.Z. Dude, this is **hella** cool!: the syntax and semantics of ‘hella’. Fall 2007 Meeting of the American Dialect Society (ADS), at the 49th Annual Midwest Modern Language Association (M/MLA) Convention, Cleveland, OH, Nov. 8-11, 2007.

Alexander, J.A., Bradlow, A.R., Ashley, R.D., and Wong, P.C.M. On the interaction between speech-pitch processing and music-pitch processing. Language and Music as Cognitive Systems, Cambridge, U.K., May 11-13, 2007.

Alexander, J.A., Wong, P.C.M., and Bradlow, A.R. Lexical tone perception in musicians and non-musicians. Northwestern Institute on Complex Systems Complexity Conference. Evanston, IL, April 20-21, 2006.

Research Experience

Su04, Sp07	Research Assistant to Ann Bradlow, Speech Communication Laboratory, Northwestern University
2002-2003	Research Associate under Jane Edwards, International Computer Science Institute, Berkeley, CA
2001-2002	Research Assistant to Ian Maddieson, Department of Linguistics, University of California, Berkeley
2000-2001	Research Assistant to Larry Hyman, Department of Linguistics, University of California, Berkeley

Teaching Experience (all at Northwestern University unless noted otherwise)

Instructor:

F06, F09, F10	LING 380, *Academic Speaking & Fluency for Non-native Speakers: American English Pronunciation*
W09	LING 380, *Academic Speaking & Fluency for Non-native Speakers: Conversation and Fluency*
Sp05	LING 380, *Academic Speaking & Fluency for Non-native Speakers: Academic Presentations*
Sp10	LING 381, *Academic Writing for Non-native Speakers*
Su07, 08, 09, 10	NU International Summer Institute (ISI), *American English Pronunciation*

Su07, 08, 09, 10	NU ISI, *American English Conversation*
Su05	NU ISI, *Test of Spoken English Preparation Course* (official title: Tutor)

Teaching Assistant:

W07, F07, W10	LING 250, *Sound Patterns in Human Language*
W05, W08	LING 222, *Language, Politics, & Identity*

Mentor:

2003-2004	NU Department of Communication Sciences and Disorders, Mentor to undergraduate Independent Study student Krupa Shah

ESL Tutor:

F04, Sp08, F08, Sp09	NU ESL program
Su07, 08, & 09	NU ISI
F04-present	Private Tutor

Grader:

2002-2003	LING 155, *Native America Meets the Europeans* (UC Berkeley)

Grants, Awards, and Honors (all at Northwestern University unless noted otherwise)

2011-2013	National Science Foundation International Research Fellowship Program award (Office of International Science and Engineering), Jennifer A. Alexander, P.I. Project title: *Acoustic Perceptual Properties of Suprasegmental Contrast Systems*. Award amount: $138,207.
2008-2009	Northwestern University Graduate Research Grant
2008	Acoustical Society of America Student Travel Grant
2005-2006	Cognitive Science Program Advanced Graduate Student Fellowship/Travel Grant
	Alternate for Helen R. Piros Weinberg Dissertation and Research Fellowship
	NU Institute on Complex Systems Complexity Conference Student Poster Award
2003-2004	University Fellowship
2003	Hellenic Times Scholarship Fund Scholarship
2002	Jesse Sawyer Fund for Applied Linguistics Award (UC Berkeley)
2002	Summer Undergraduate Research Fellowship (UC Berkeley)

Professional Service (all at Northwestern University unless noted otherwise)

2010	Linguistics Department Transition Team member
2008	Ad-hoc reviewer for the Journal of the Acoustical Society of America
05-06; 07-10	Linguistics Department Video Coordinator
2007-2008	Language and Music Systems Co-Liaison
2006-2007	Linguistics Department Social Chair
2004	Organizing Committee, Mid-Continental Workshop on Phonology
2003-2005	Linguistics Department Receptions Committee