TONE-TUNE ASSOCIATION AND VOICE QUALITY IN GREEN MONG FOLK SONGS

Anna Mai, Mai Moua, and Marc Garellek

University of California, San Diego
acmai@ucsd.edu, mmaimoua@gmail.com, mgarellek@ucsd.edu

ABSTRACT

In this paper we explore the relationship between lexical tone and musical tune in folk songs of Green Mong, a Hmong-Mien language spoken in Southeast Asia and by Hmong diaspora communities. Previous research has claimed that, in folk songs, the seven lexical tones of the language map onto four musical notes. Non-modal quality is impressionistically neutralized in the breathy and creaky tones when sung. We test these claims quantitatively using a corpus of Green Mong folk songs sung from memory by four speakers (two men and two women). The songs were transcribed, and vowels were acoustically segmented and annotated for musical phrasing, and were analysed for pitch and voice quality. Results indicate that, in accordance with previous claims, the seven tones map onto four pitch targets, but this varies by phrasing as well as by singer. Additionally, we find some evidence that voice quality differences do remain for the non-modal tones in Mong.

Keywords: text-setting, tone, phonation, Hmong-Mien

1. INTRODUCTION

Green Mong has seven lexical tones distinguished both by pitch and voice quality [3, 14]. The goal of this paper is to determine how the seven tones map onto musical notes in Green Mong song. Previous work claims (1) that these seven tones map onto four musical notes, and (2) that voice quality distinctions are neutralized in folk songs known as *kws txiaj* (in White Hmong) [6], or *lug txaj* in Green Mong. In this paper we test these claims for Green Mong using a corpus of songs (including both *lug txaj* and other genres) sung from memory by four speakers (two women, two men) and examine the realization of tone pitch and voice quality across song types.

Green Mong (also known as Mong Leng or Mong Njua) is a Hmong-Mien language of the West Hmongic branch spoken by over a million speakers in southern China, Southeast Asia, and in Hmong diaspora communities [3]. It is closely related to White Hmong (also known as Hmong Daw). Each syllable of Green Mong bears one of seven lexically contrastive tones. In the standard Romanized Popular Alphabet (RPA) orthography, tone identity is indicated by the final letter of each syllable, and we use these letters to refer to tones throughout. Tone contours are high (*-b*), high falling (*-j*), mid (unmarked), mid falling and breathy (*-g*), mid rising (*-v*), low (*-s*), and low falling creaky (*-m*) [3, 2, 1]. A summary of tone contours and qualities is shown in Table 1.

Table 1: Summary of Green Mong tonal system and correspondences between tones and musical notes. Column 1 indicates the orthographic letter associated with each tone. The mid tone is not orthographically marked in RPA. Column 2 gives the description associated with each tone in [1]. Column 3 provides the description and approximate semitone value of the musical note which corresponds to each tone in [6].

Tone (RPA)	Description (Andruski 2006)	Musical note (Catlin 1997)
-b	High	Highest (10)
-j	High-falling	Lowest (0)
-g	Mid-falling breathy	Mid-low (5)
∅	Mid	Mid-high (7)
-v	Mid-rising	Highest (10)
-s	Low	Lowest (0)
-m	Low-falling creaky	Mid-low (5)

Lug txaj are traditional folk songs typically sung during courtship and the Hmong New Year. [6] describes the scale system of these folk songs as tetratonic, with four notes per octave. Throughout this paper, these notes are referred to by the number corresponding to their chromatic semitone value above the lowest note in the song (0–12), where 0 and 12 are octave-equivalent.

Catlin [6] speculates that Hmong tones (both for White Hmong and Green Mong) may map on to the four notes of *lug txaj* in accordance with an archaic tonal system, as has been argued for the areal music traditions of Sino-Tibetan and Austro-Thai groups

[5]. As reconstructed in [15, 16], the tonal system of Proto-Hmong-Mien comprised four tones, labeled A, B, C, and D. The seven tones of the present-day system may have initially developed from a segmentally-conditioned split of the four tones into two registers, where each of the tones developed two variants conditioned on the voicing of the syllable-initial consonant.

In this paper, we investigate the mapping between the musical notes of *lug txaj* and the tones of present-day Green Mong and quantitatively assess (1) the degree to which they correspond to an older tonal system and (2) the degree to which characteristic voice quality is maintained during song.

2. METHODS

2.1. Speakers

Data in this paper were collected from four speakers of Green Mong from the Xieng Khouang province of Laos (two women, two men). All speakers were over the age of 50 and had been living in the United States for more than 20 years at the time of recording. They were all proficient in both Green Mong (their native variety) and White Hmong, as well as English.

2.2. Recording of the folk songs

Monophonic recordings were made in a sound-attenuated booth collected using Praat [4] software and a preamplified head-mounted microphone. Speakers sang freely from memory for as long as they felt comfortable singing, and following recording, songs were transcribed by the second author, a native speaker of Hmong fluent in both Green Mong and White Hmong. Songs were then segmented in Praat by phonetically-trained undergraduate research assistants.

The songs in this analysis differ by type and regional style. The two women (FM1 and FK1) both sang *lug txaj* courtship songs, though they differed in style. FM1 sang *lug txaj Moob moos pheeb*, a *lug txaj* of the style from the village *Moos Pheeb*. FK1 sang *lug txaj Moob xeev*, a *lug txaj* of the style found throughout the Xieng Khouang province of Laos.

One of the male speakers, MY1, sang two *lug txaj* songs, one a courtship song and the other a New Year song (*lug txaj xyoo tshab*). He also sang two *zaaj tsoob* songs typically sung at weddings. The other male speaker, MV1, first played the *qeej* bamboo reed instrument, and then (unaccompanied by the *qeej*) sang the words that are conveyed by the instrument. This style of singing is known as *pav zaaj qeej* (in White Hmong *piav zaj qeej*) or 'explaining the *qeej* music.' In total, the four speakers recorded 14 songs (=4,897 syllables): 10 *lug txaj* courtship songs, two *zaaj tsoob* wedding songs, and two *pav zaaj qeej* songs.

2.3. Word and vowel annotation

Recordings were segmented by word and by vowel, and were labeled for lexical tone and utterance position. Initial boundaries for words beginning with aspirated or unaspirated stops were placed at the beginning of stop closure, assessed visually as a brief period of silence in the spectrogram. Initial boundaries for words beginning with prenasalized stops were also placed at the beginning of stop closure, assessed visually as a period of high-frequency silence in the spectrogram. Initial boundaries for words beginning with fricatives were placed at the beginning of frication, assessed visually as noise in the spectrogram. And initial boundaries for words beginning with sonorant consonants were placed at the beginning of the drop in intensity preceding the vowel.

The criteria for word-final boundaries varied by phrase position. For words in initial or medial phrase positions, the final word boundary was taken to be the initial boundary of the next word. Because every word in Mong ends with a vowel or nasal coda, the word-final boundary for phrase-final words was placed at the offset of the second oral formant if vowel-final or at the offset of the first nasal formant if nasal-final.

For vowel segmentation, initial and final boundaries were determined by the onset and offset of the second formant, respectively.

In addition to word, vowel, and tone labels, each word was also labeled for utterance position (Initial, Medial, Final), which was determined by whether a breath occurred before or after the word. If no breath could be heard, pauses of >300ms were also considered utterance-final. Words that occurred as whole utterances (i.e., ones that are both utterance-initial and -final) were excluded.

2.4. Acoustic analysis

Segmented vowels were then analysed acoustically using VoiceSauce [17], which outputs values for a variety of acoustic measures every millisecond of a labelled interval. In this study, we focus on two acoustic measures: F0, calculated using the STRAIGHT algorithm [11], and Cepstral Peak Prominence (CPP) [10], a harmonics-to-noise ratio measure.

The output measures were averaged over the entire vowel, because unlike in spoken Mong, sung

Mong vowels have temporally-stable characteristics. For each speaker, we converted the F0 values to semitones, using the speaker's average F0 as the baseline. We then normalized F0 values by tone and removed outliers greater than 2.5 standard deviations from the mean.

Cepstral Peak Prominence (CPP) was used to compare the voice quality between the mid-falling breathy *-g* tone and its non-breathy (modal) counterpart (which in song is the mid tone, as discussed below), and between the low-falling creaky *-m* tone and its non-creaky counterpart, the low modal *-s* tone. Non-modal (both breathy and creaky) voice qualities are typically noisier than modal voice, with lower values of harmonics-to-noise ratio measures [9, 12, 13, 7]. Consequently, we expect that, should voice quality differences remain in sung Mong, the mid-falling breathy tone and the low-falling creaky tone will be noisier (i.e., will have lower CPP values) than their modal counterparts.

3. RESULTS

3.1. Tone-tune association

The scale type of each of the 14 songs in the dataset was identified as either tetratonic or pentatonic. The majority of songs (10 of the 14) are tetratonic, and the most common scale is 0-5-7-10, where each number refers to the semitone difference above the lowest note 0, as described by [6]. Given the size of the current corpus, it appears possible that some differences in scale may be due to variation across speakers. For example, speakers FM1 and MV1 use only one scale each, and it is not clear whether this consistency is due to speaker preference or to the particular identity of the songs that were sung. However, speakers FK1 and MY1 each produced multiple scales, with MY1 producing four songs with four distinct scales. These results suggest that music in this tradition uses multiple tetratonic and pentatonic scales. A summary of the musical scales by song and speaker is shown in Table 2.

Next we look at how each lexical tone mapped on to these notes. Because the scale varied by song, we divided the notes up according to six values: 0 (the lowest note in the song), 2, 4/5 (depending on whether the song's scale had a note 4 or 5 semitones above the lowest), 7/8 (depending on whether the song's scale had a note 7 or 8 semitones above the lowest), 9/10 (depending on whether the song's scale had a note 9 or 10 semitones above the lowest), and 12. Note that every song had notes labeled '12'; that is, one octave higher than the lowest. Our results are largely in accordance with those of [6]; however, there are some notable differences. In particular, all tones can appear with any note. Moreover, the low *-s* tone in our corpus is generally assigned to a mid-low note, rather than to the lowest one; see Table 3.

Table 2: Summary of the scales used in songs in the dataset.

Singer/Song	Description	Scale
FK1/1-2	Courtship song	0-5-7-10
FK1/3-4	Courtship song	0-2-5-7-10
FK1/5-6	Courtship song	0-5-7-10
FM1/1-2	Courtship song	0-2-5-7-9
MV1/1-2	*Qeej* song	0-4-7-9
MY1/1	Wedding song	0-2-7-9
MY1/2	Courtship song	0-4-7-9
MY1/3	Wedding song	0-5-7-10
MY1/4	New Year song	0-5-8-10

Table 3: Summary of the scales used in songs in the dataset, compared to analysis in [6]. The third column shows what percentage of tokens per tone were realized as a given note. The bolded numbers refer to the two most frequent notes assigned to each tone.

Tone	Musical note	Corpus %					
	Catlin (1997)	0	2	4/5	7/8	9/10	12
High *-b*	10	13	5	15	20	**21**	**27**
High-falling *-j*	0	**26**	11	**19**	17	16	11
Mid-falling *-g*	5	11	10	**24**	**25**	20	10
Mid	7	11	5	16	**29**	**27**	12
Mid-rising *-v*	10	10	7	**23**	21	**23**	16
Low *-s*	0	17	11	**21**	**26**	17	8
Low-falling *-m*	5	13	12	**23**	**25**	17	10

3.2. Effects of musical phrasing

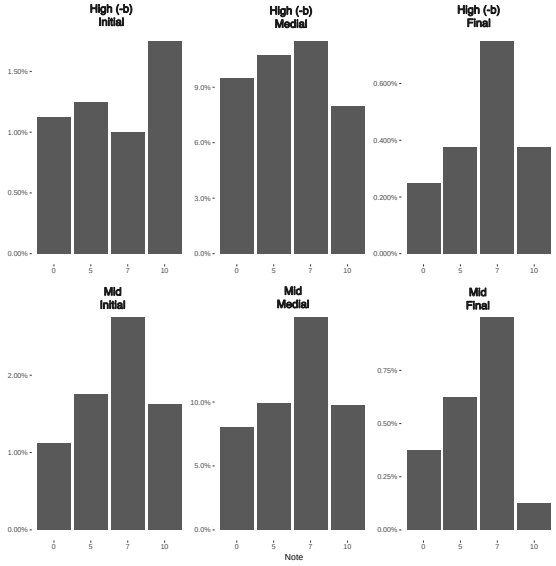
Given that each word was coded for its position within the musical utterance, we can determine whether the tone-tune alignment differs by phrasal position.

Tones can appear with any note in all phrasal positions, but some tones show more variation by phrasal position than others. For instance, in songs of scale 0-5-7-10, the mid tone most often occurs with note 7 (as predicted by [6]), regardless of phrasal position. On the other hand, the high *-b* tone, predicted to be note 10, is most often that level (rather than other levels) in utterance-initial position; in medial and final positions, it most frequently occurs on note 7; see Figure 1.

3.3. Voice quality

We also tested whether non-modal voice quality is retained in song, despite claims to the contrary [6].

Figure 1: Distribution of notes for the high *-b* and mid tone by phrasal position, for songs sung on a 0-5-7-10 scale. The y-axis represents the percentage of time each bar is represented in the entire corpus.



That is, are the mid-falling breathy *-g* tone and the low-falling creaky *-m* tone noisier than their modal counterparts? For each speaker, we ran two linear mixed-effects models on phrase-medial vowels: one modeling CPP as a function of Tone ('breathy' *-g* vs. modal mid), the Musical Note, and their interaction; the other model similarly looked at CPP as a function of Tone ('creaky' *m* vs. modal *-s*), the Musical Note, and their interaction. The mid-falling breathy *-g* tone was compared to the mid tone because they have similar distributions across notes. For both models, we included random intercepts by Word and a maximal random-effects structure. We excluded non-medial vowels due to their relative infrequence in the corpus.

For most speakers, there was no significant main effect of Tone (no Tone*Note interactions were significant), suggesting no difference in voice quality. However, one speaker, FK1, showed lower CPP values for the breathy tone compared to its non-breathy counterpart ($\beta = 0.81$, $t = 2.22$, $p < 0.05$).

4. DISCUSSION AND CONCLUSION

In this paper, we investigated the mapping of linguistic tones to musical notes in traditional Green Mong song, and quantitative results from this corpus confirm some previously reported findings and refute others. In particular, songs in this corpus confirm the generalization that scales in this musical tradition are predominantly tetratonic, as reported in [6]. However, in this corpus, we observe six distinct scales, both tetratonic and pentatonic. Many different tetratonic and pentatonic scales exist, but those observed in this corpus appear to reflect a principled subset of those scale types. All scales in the corpus are *anhemitonic*, meaning that they contain no intervals of a single semitone. Additionally, several interval types are common across scales, including the major second, minor third, and perfect fourth.

Furthermore, the mapping of linguistic tones onto musical notes in these songs resembles the mapping given in [6], as shown in Table 3, but is also far more variable than previously reported. While previous reports indicate that the mapping between linguistic tones and musical notes is consistent, we find that linguistic tones are spread more evenly across notes of the musical scale. Of the seven linguistic tones, the note previously ascribed to each tone is among the two most frequent notes assigned for six of the tones. This finding substantially weakens the claim in [6], but confirms its basic generalization.

However, contrary to claims in [6], we find that at least one speaker maintains a voice quality contrast during song. While none of the speakers maintained the characteristic quality of the creaky *m*-tone, breathy *g*-tones in songs sung by FK1 had significantly lower CPP than the closest corresponding modal tone. In a perceptual study assessing acoustic cues to tone identity in closely-related White Hmong, listeners used breathy voice but not creaky voice to cue tone identity [8]. It's possible that this asymmetry in the use of voice quality cues explains the lack of characteristic voice quality distinction on creaky *m*-tones, but given that only one speaker in our corpus appears to maintain the characteristic quality of the breathy *g*-tone during song, a larger corpus of speakers would be required to fully assess this possibility.

Finally, we believe there is truth to the claim by [6] that the tones of *lug txaj* reflect an older stage of tone production in the Hmong languages. In the tonal reconstruction in [15], all seven tones derived from an older two-register system; tones usually sung on the highest notes (high *-b*, mid, and mid-rising *-v*) belonged to the high-pitched register, and the remaining tones, which tend to be sung on lower notes, belonged to the low-pitched register. Still, given our finding that Green Mong tones and notes do not have a one-to-one mapping, it is likely that what was once a stricter register-to-musical note correspondence has yielded in the present day to a looser mapping between tones and melody.

5. REFERENCES

[1] Andruski, J. E. 2006. Tone clarity in mixed pitch/phonation-type tones. *Journal of Phonetics* 34, 388–404.

[2] Andruski, J. E., Costello, J. 2004. Using polynomial equations to model pitch contour shape in lexical tones: an example from Green Mong. *Journal of the International Phonetic Association* 34, 125–140.

[3] Andruski, J. E., Ratliff, M. 2000. Phonation types in production of phonological tone: the case of Green Mong. *Journal of the International Phonetic Association* 30, 37–61.

[4] Boersma, P., Weenink, D. 2015. Praat: doing phonetics by computer [Computer program]. Version 5.4.04. Retrieved January 5, 2015 from http://www.praat.org/.

[5] Bradley, D. 1979. Speech through music: The Sino-Tibetan gourd reed organ. *Bulletin of the School of Oriental and African Studies* 42, 535–540.

[6] Catlin, A. 1997. Puzzling the text: Thought-songs, secret languages, and archaic tones in Hmong music. *The World of Music* 39, 69–81.

[7] Garellek, M. to appear. The phonetics of voice. In: Katz, W., Assmann, P., (eds), *Routledge Handbook of Phonetics*. Routledge.

[8] Garellek, M., Keating, P., Esposito, C. M., Kreiman, J. 2013. Voice quality and tone identification in White Hmong. *Journal of the Acoustical Society of America* 133, 1078–1089.

[9] Gordon, M., Ladefoged, P. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29, 383–406.

[10] Hillenbrand, J., Cleveland, R. A., Erickson, R. L. 1994. Acoustic correlates of breathy voice quality. *Journal of Speech and Hearing Research* 37, 769–778.

[11] Kawahara, H., de Cheveigné, A., Patterson, R. D. 1998. An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: revised TEMPO in the STRAIGHT-suite. *ICSLP-1998* 0659.

[12] Keating, P., Esposito, C., Garellek, M., Khan, S., Kuang, J. 2011. Phonation contrasts across languages. *Proceedings of the International Congress of Phonetic Sciences* Hong Kong. 1046–1049.

[13] Keating, P., Garellek, M., Kreiman, J. 2015. Acoustic properties of different kinds of creaky voice. *Proceedings of the 18th International Congress of Phonetic Sciences* Glasgow.

[14] Ratliff, M. 1992. *Meaningful Tone: A study of tonal morphology in compounds, form classes and expressive phrases in White Hmong*. Monograph Series on Southeast Asia. DeKalb, IL: Northern Illinois University, Center for Southeast Asian Studies.

[15] Ratliff, M. 2010. *Hmong-Mien language history*. Australian National University.

[16] Ratliff, M. 2015. Tonoexodus, tonogenesis, and tone change. In: Honeybone, P., Salmons, J., (eds), *The Oxford Handbook of Historical Phonology*. Oxford: Oxford University Press 245–261.

[17] Shue, Y.-L., Keating, P. A., Vicenik, C., Yu, K. 2011. VoiceSauce: A program for voice analysis. *Proceedings of the International Congress of Phonetic Sciences* Hong Kong. 1846–1849.