TONE PERCEPTION IN SGAW KAREN

Marc Brunelle & Joshua Finkeldey

University of Ottawa, Canada
marc.brunelle@uottawa.ca; jfink082@uOttawa.ca

ABSTRACT

Sgaw Karen has a complex tone system based on voice quality, f0 and duration. A perceptual study reveals that voice quality and duration are central to its patterns of tone contrast. F0-based identification cues are the presence of a low offset and of a falling contour. No evidence was found for the decompositionality of contours.

Keywords: Sgaw Karen, tone, voice quality, perception

1. INTRODUCTION

Southeast Asian languages often have large tone inventories that combine complex pitch contours with voice quality distinctions [1, 3, 6, 8, 10]. However, it is still unclear to what extent these pitch and voice quality properties are perceptually redundant. Moreover, studies of tone perception in Southeast Asian languages have brought forward conflicting evidence about the phonetic properties used for tone identification and tone representation. Tone contours have been argued to be unitary in Vietnamese [3, 8], but perceptually and phonologically decomposable in Thai [11, 12]. These differences could well be due to language specific properties: it is impossible to make cross-linguistic generalizations due to lack of evidence.

Sgaw Karen is a Tibeto-Burman language of Burma that has a typical Southeast Asian tone system composed of 6 complex contour tones that can be breathy or creaky. The tones of Karen have mostly been studied from a diachronic perspective [4, 5, 7]. Most descriptions of Sgaw Karen phonology also include a section on tone [2, 9], but to our knowledge it has never been studied instrumentally.

The goals of this study are to describe the acoustic properties of Sgaw Karen tones (§2) and the phonetic properties used for tone identification (§3). These results will shed light on the types of phonetic cues that used across languages and will bring further evidence for the phonological structure of tone contrasts.

2. ACOUSTIC PROPERTIES OF SGAW KAREN TONES

2.1. Methods

Four speakers of Sgaw Karen (2 men, 2 women) born between 1982 and 1991 in the Karen State or in refugee camps in Thailand recorded the 24 words created by combining the syllables /na, da, ta, t^{h}a/ with the six tones in the carrier sentence in (1). Each word was read 6 times.

(1) jəp^{h}a^4 li^{3}mɛ^{3}p^{h}lɤ^5 ____ lɤ^1 mɛ^{2}k^{h}ro^{5}fo^5 pu^6
I-read word at microphone inside
I read the word ____ into the microphone.

The f0 and H1-H2 of target vowels were measured at their onset, offset and at 8 additional equidistant points. Vowel duration was also measured. All measurements were made in *Praat 5.1.26*.

2.2. Results

Results from a representative speaker (M1) are reported in Figures 1-3 (*cf. audio files 1-6*).

Figure 1: Mean H1-H2 of the 6 tones, speaker M1.

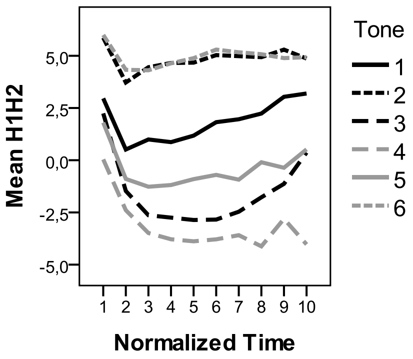
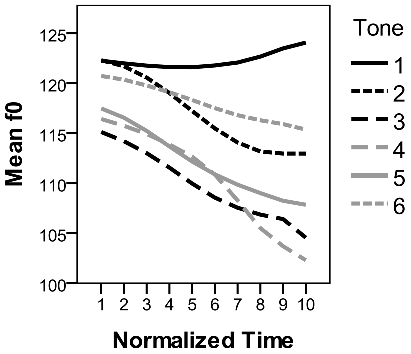


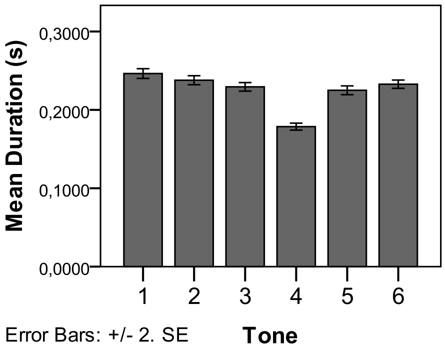
Fig. 1 illustrates the salience of voice quality in Sgaw Karen. While tones 1 and 5 fall in the mid H1-H2 range (full lines), which suggests a modal voice quality, tones 2 and 6 have a high H1-H2, indicative of breathiness or laxness (dense broken lines), and tones 3 and 4 have a low H1-H2, indicative of creakiness or tenseness (sparse broken lines). Tone 4 often ends in a clear glottal stop. The other male speaker has a similar distribution, while the two female speakers do not distinguish tones 1, 2, 5 and 6 as clearly.

Figure 2: Mean f0 of the 6 tones, speaker M1.



The f0 distribution in Fig. 2 contrasts a slightly rising tone 1 with the 5 other tones, which are all moderately falling. Subject F2 has similar contours, while the tones of subjects M2 and F1 have a similar relative distribution, but with a counterclockwise rotation that makes the falling tones level or even slightly rising. This general change in contour is likely due to the different intonation used in reading the wordlist. Another important point is that although tone 1 always ends highest, the relative height of the other tones is highly variable across subjects.

Figure 3: Mean duration of the 6 tones, speaker M1.



Duration results (Fig.3) are consistent across subjects. Tone 4 is consistently shorter than the other 5 tones, which are all of comparable duration.

3. PERCEPTION EXPERIMENT

3.1. Methods

3.1.1. Stimuli and procedure

Natural and resynthesized stimuli were used in this experiment. Resynthesis was carried out in *Praat 5.1.26*. *Natural* stimuli are used as a control group consisting of one representative natural utterance of each of the six tones on the syllable [na], produced in isolation by a male speaker (*cf. audio files 7-12*). *Duration-manipulated* stimuli are resynthesized versions of the natural stimuli of 4 tones chosen for their characteristic voice qualities (T1: modal, T2: breathy, T3: creaky, T4: glottalized). The four tones were rescaled to identical durations (onset: 150 ms, vowel: 400 ms) and tones 1 and 3 were shortened (onset: 150 ms, vowel: 146 ms) to make them comparable to the natural utterance of tone 4 (*cf. audio files 13-21*). *Pitch-manipulated* stimuli were resynthesized by superimposing twenty-five f0 contours composed of all the combinations of onsets and offsets set at 5 f0 levels (100, 115, 130, 145 and 160 Hz) over the vowels of the 6 duration-manipulated stimuli. The wide f0 range corresponds to values observed in citation forms (contrast with values in carrier sentence in §2). In total, 162 stimuli were tested (6 natural, 6 duration-manipulated, 150 pitch-manipulated).

The stimuli were played to native speakers in an identification experiment run with *Presentation 14.8*. Participants had to identify them by clicking on one of six response buttons on which the syllable [na] with the six tones was written in Karen script. Short instructions were given at the top of the screen in both Karen and English. The 162 stimuli were played 5 times each, in 5 different blocks. The order of presentation was randomized independently in each of the 5 blocks. Responses and reaction times were measured.

3.1.2. Participants

Nineteen natives speakers of Sgaw Karen, born between 1940 and 1993, participated in the experiment (11 men, 8 women). Seven of them had to be excluded, either because of unfamiliarity with computers and experimental tasks (2) or because they failed to identify natural tones above chance (5), leaving 12 participants. They were all raised in the Karen State, Burma, or in refugee camps in Thailand. Although they are all dominant in Karen, many of them also speak Burmese or Thai and they all had at least basic English skills. Participants had been living in Ottawa for 1-7 years at the time of the experiment.

3.2. Results

3.2.1. Natural stimuli

Natural stimuli were correctly identified well-above chance, but with significant error patterns (Table 1). Tones are often confused because of their voice quality: tones 2 and 6 share a high H1-H2 (breathiness), while tones 3 and 4 both have low H1-H2 (creakiness). The fact that none of our participants could properly identify tone 5 will be discussed in §4.

Table 1: Response matrix for natural stimuli (60 stimuli /tone).

	Responses						
Stimuli	T1	T2	T3	T4	T5	T6	Ø
Tone 1	**47**	3	1	2	5	1	1
Tone 2	1	*24*	5	1	0	**29**	0
Tone 3	0	4	**29**	16	8	3	0
Tone 4	0	2	23	**28**	6	0	1
Tone 5	2	10	15	10	*6*	**17**	0
Tone 6	0	20	3	2	1	**34**	0

3.2.2. *Duration-manipulated stimuli*

Duration-manipulated stimuli confirm the results obtained with natural stimuli, but show that duration affects identification (Table 2). Shortened versions of tones 1 and 3 mostly elicit tone 4 responses (note the high proportion of tone 5 responses for shortened tone 1). The duration-normalized versions of tones 1, 2, and 3 obtain the same rough response distributions as their natural correspondents, as expected. Note however that the responses elicited by the lengthened version of tone 4 are not very different from its natural version, suggesting that while shortness is a strong cue, a long duration might be less relevant.

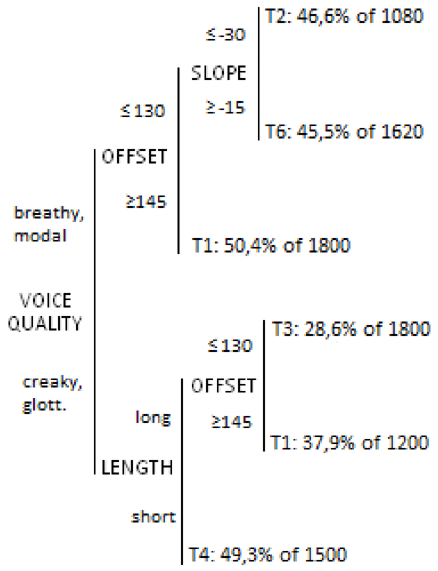
Table 2: Response matrix for duration-manipulated stimuli (60 stimuli /tone).

		Responses						
Length	Stimuli	T1	T2	T3	T4	T5	T6	Ø
Short	Modal (T1)	12	2	5	**21**	20	0	0
	Creaky (T3)	1	1	23	**27**	6	0	2
Long	Modal (T1)	**47**	6	0	0	1	4	2
	Breathy (T2)	2	21	1	1	1	**34**	0
	Creaky (T3)	0	4	**25**	12	10	6	3
	Glott. (T4)	1	8	20	**22**	4	5	0

3.2.3. *Pitch-manipulated stimuli*

A binary classification and regression tree (CRT) of the responses given by participants is plotted in Fig. 4. The dependant variable is the *response* given by participants and the predictor variables are *voice quality* (breathy, modal, creaky or glottalized), *length* (long or short), *f0 onset* and *f0 offset* (100, 115, 130, 145 or 160 Hz) and *slope*, as calculated by subtracting the onset from the offset. New branches were created if the Gini impurity score of terminal nodes could be increased by at least 0,01. The tree was pruned at a maximum difference in risk of 1 standard error. Terminal branches give the response that was the most frequently given, its prevalence and the total number of tokens in this branch. The second most frequent responses never reach 25% in any given terminal node.

Figure 4: Classification tree of responses given for pitch-manipulated stimuli (n = 9000).



The highest split in Fig. 4 is *voice quality*, which means that modal and breathy stimuli are treated differently from creaky or glottalized ones. Breathy and modal stimuli are then separated by *f0 offset*: a high f0 (≥145 Hz) yields 50,4% of tone 1 responses, while low offset stimuli (≤ 130 Hz) are further split by *slope*. A clearly falling slope (≤ -30 Hz) mostly elicits tone 2 responses (46,6%), but tone 6 is the preferred response (45,5%) for non-falling or slightly falling slope (≥ -15 Hz).

Creaky or glottalized tokens, on the other hand, are distinguished by *length*. Short tokens are mostly identified as tone 4 (49,3%), while long tokens are further divided by *f0 offset*. Once again, a high f0 (≥145 Hz) is a cue to tone 1 (37,9%). In this case though, a low offset stimulus (≤ 130 Hz) is mostly identified as tone 3, but only 28,6% of the time.

4. DISCUSSION

Inter-speaker variation found in the acoustic results already suggest that some phonetic properties are associated with Sgaw Karen tones in a more stable way than others. First, all 4 speakers maintain a clear H1-H2 contrast between tones 3 and 4 and the other tones, while tones 1, 2, 5 and 6 are not always clearly distinguishable. Second, the only tone that has a distinct contour is tone 1, which is systematically rising. Other tones are all preferably falling, but this contour seems easily overridden by intonation (speakers M2 and F1). The relative height of the 6 tones also seems variable across speakers, although tone 1 consistently has a higher offset than other tones.

Identification results confirm these observations. First, participants do not use all voice quality nuances for identification. They only contrast breathy/modal with creaky/glottalized stimuli. Second, vowel length is only used in interaction with voice quality: short and creaky/glottalized stimuli are identified as tone 4, but length is not relevant in breathy/modal stimuli. Third, tone height plays a limited role: stimuli with high offsets (145 Hz or above) are systematically identified as tone 1, unless they are short *and* creaky/glottalized. Finally, slope is used to distinguish breathy/modal stimuli with a low offset (130 Hz or below). They are identified as tone 2 if they have a clearly falling slope; otherwise they prompt tone 6 responses. Tone 3 seems to be a relatively ill-defined category: stimuli that are creaky/glottalized, long and have low offsets tend to elicit this answer, but not decisively.

The most unexpected result in this study is the fact that tone 5 is rarely given as a response and is not associated with well-defined acoustic properties. In fact, significant proportions of tone 5 responses are only obtained for stimuli that are acoustically different from all natural tones (like the shortened version of tone 1 in Table 2). Post-experiment discussions with speakers lead us to believe that this tone might have merged with another tone in Sgaw Karen, while being preserved in the normative language and script, leading to some confusion.

These results suggest similarities between Sgaw Karen and Northern Vietnamese tones. Voice quality is crucial in both tone systems, but its contrastive role seems limited to an opposition between modal/breathy on the one hand and creaky/glottalized on the other hand [3, 8]. Moreover, the fact that slope distinguishes tones 2 and 6, along with the fact that f0 height is used for identification at tone offset, but not at tone onset, argues against the decomposition of contours into sequences of level tones, as in Vietnamese [3, 8].

On the other hand, there are also important differences between Vietnamese and Sgaw Karen. The only relevant contour opposition in Sgaw is between falling and non-falling, while Vietnamese is sensitive to a rising/non-rising opposition [3, 8]. Furthermore, f0 offset is only used to distinguish high offset tone 1 from non-high tones, whereas Vietnamese is organized around a low/non-low offset opposition [3]. Another important difference is that tone duration, in interaction with voice quality, is an important identification cue in Sgaw Karen.

Our results confirm that Sgaw Karen, like many Southeast Asian languages, has a tone system that is heavily based on non-pitch properties. The role of f0 height and f0 contour in its patterns of tone contrast is limited. However, important differences in the exact thresholds used to define relevant categories of contours and tone offsets, along with an absence of evidence for the decompositionality of tone contours, suggest that the phonetic properties used as tonal identification cues are largely language-specific.

5. ACKNOWLEDGEMENTS

Special thanks to Nimrod Andrew, our main consultant, who also helped us recruiting participants. This study would not have been possible without him.

6. REFERENCES

[1] Andruski, J.E., Ratliff, M. 2000. Phonation types in production of phonological tone: the case of Green Mong. *J. of the Int. Phonetic Association* 30, 37.

[2] Baa, S.L. 2001. *The phonological Basis of a Sgaw and Northwest Karenic Orthography*. MA. Chiang Mai: Payap Un.

[3] Brunelle, M. 2009. Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics* 37, 79.

[4] Haudricourt, A-G. 1953. A propos de la reconstitution du karen commun. *Bull. Soc. de Ling. de Paris* 49, 129.

[5] Haudricourt, A.-G. 1975. Le système de ton du karen commun. *Bull. de la Soc. de Ling. de Paris* 70, 339.

[6] Huffman, M.K. 1987. Measures of phonation type in Hmong. *J. of the Acoustical Society of America* 81, 495.

[7] Jones, R.B. 1961. Laryngeals and the development of tones in Karen. *Burma Research Soc. Fiftieth Ann. Publication*. Rangoon: Burma Res. Soc. 101.

[8] Kirby, J. 2010. Dialect experience in Vietnamese tone perception. *J. of the Acoustical Society of America* 127.

[9] Manson, K. 2005. *Tone Patterns of Karen Languages*. Manuscript, Chiang Mai: Payap University.

[10] Michaud, A. 2004. Final consonants and glottalization: new perspectives from Hanoi Vietnamese. *Phonetica* 61, 119.

[11] Morén, B., Zsiga, L. 2001. Markedness and lexical tone in standard Thai: Phonetics and phonology. *Proc. of the 27th Annual Meeting of the Berkeley Linguistics Society*.

[12] Zsiga, E., Nitisaroj, R. 2007. Tone features, tone perception, and peak alignment in Thai. *Language and Speech* 50, 343.