

# Perceptual Cues to Lexical Tone in Burmese

James Gruber  
Georgetown University

## 1. Background

This study tested the perception of multiple phonetic cues to lexical tone in Burmese via a forced-choice identification task using natural and resynthesized stimuli. Tone in Burmese is a complex contrast where systematic differences in intensity, duration, phonation type, and vowel quality operate alongside pitch distinctions. Prior phonetic studies, often attempting to establish whether pitch or phonation properties primarily signal the contrast, have confirmed that a distinction for each of these phonetic qualities is reliably produced by speakers (Thein Tun 1982, Javkin and Maddieson 1983, Watkins 2005, Gruber 2010). The table in (1) demonstrates these phonetic qualities.

(1)	Phonetic Descriptions (Bradley 1982, Wheatley 1987, Watkins 2001)				
<u>Tone</u>	<u>Duration</u>	<u>Intensity</u>	<u>Pitch</u>	<u>Phonation</u>	<u>Syll. Type</u>
<i>High</i>	long	moderate	high rise or fall	breathy or plain	CV or CVN
<i>Low</i>	long, moderate	low	low level	plain, modal	CV or CVN
<i>Creaky</i>	short	high	high, sharp fall	creaky, weak glottal closure	CV or CVN
<i>Checked</i>	extremely short	highest	initial high, fall	abrupt, complete glottal closure	CVO

Non-modal voicing however, has been difficult to detect acoustically (Watkins 1997) and electroglottographic data has found only a limited two-way distinction between two more open and two more constricted tones (Watkins 2005, Gruber 2010). Thurgood (1978: 224) offers a revealing comment on the predictability of marked phonation types in Burmese, claiming that every tone “may occur with a clear voice quality, the breathy voice is found only with the heavy (High) tone, the creaky voice is found only with the creaky tone, while the level (Low) tone always occurs with the unmarked clear voice quality.” Bearing these observations in mind, the present study asks what role the multiple phonetic cues play in the perception of the four tones.

## 2. Methodology

Stimuli were composed of tokens resynthesized to controlled values for duration, phonation type, F0 height, and F0 slope, as well as a set of twenty-four natural, unaltered samples. All tokens were produced by a single speaker, a 40-year-old female with 15 years of U.S. residency. For resynthesis, 216 tokens were created from six syllables, each of two carrier syllables (/la/, /lɛ/) produced with breathy, creaky, and modal voicing. In this way, phonation type was not synthesized, but controlled for by these prototype tokens which were selected by the researcher for spectral qualities confirmed through

spectrographic analysis and an accompanying EGG trace. Other acoustic properties were modified as follows:

- **Intensity** profiles were standardized across tokens for each vowel: [la] syllables to 77dB, [lɛ] syllables to 75dB.
- **Vowel duration** was rescaled to three lengths: 175ms, 250ms, 325ms.
- **Fundamental frequency** was resynthesized to fit twelve contours:
  - 5 even pitch tracks at 20 Hz intervals from 160 Hz to 240 Hz.
  - 7 rising or falling pitch tracks with gradual or late changing slopes.

“Gradual” slopes were a straight-line interpolation from onset to offset F0 while the “late” abrupt slopes maintained the onset F0 until the 2/3<sup>rd</sup> point of the vowel before rising or falling over the final 1/3<sup>rd</sup>. All stimuli were tone-bearing monosyllables presented in isolation. The twenty-four control tokens were composed of eighteen natural citation-form productions and the six prototype phonation-bearing tokens with standardized intensity.

Ten native Burmese speakers performed the task (4 male, 6 female). All were residents of the U.S. relocated from Rangoon and were primarily between 36 and 50 years of age. One subject was 19 years old and two were graduate students in their 20s. Prior to the perception experiment, all subjects had completed a 3-hour production task (Gruber 2010).

The experiment was conducted with the DMDX software (Forster & Forster 2002) run on a Windows XP laptop. Subjects were presented a stimulus and offered a forced-choice selection between four Burmese orthographic symbols representing the /la/ or /lɛ/ syllable with each tone. If subjects desired, they were permitted to replay stimuli. Before the experiment, a brief familiarization session with sixteen natural tokens was conducted to instruct participants and set speaker volumes appropriately. The 240 stimuli were presented in six blocks of 40 tokens, each ordered randomly.

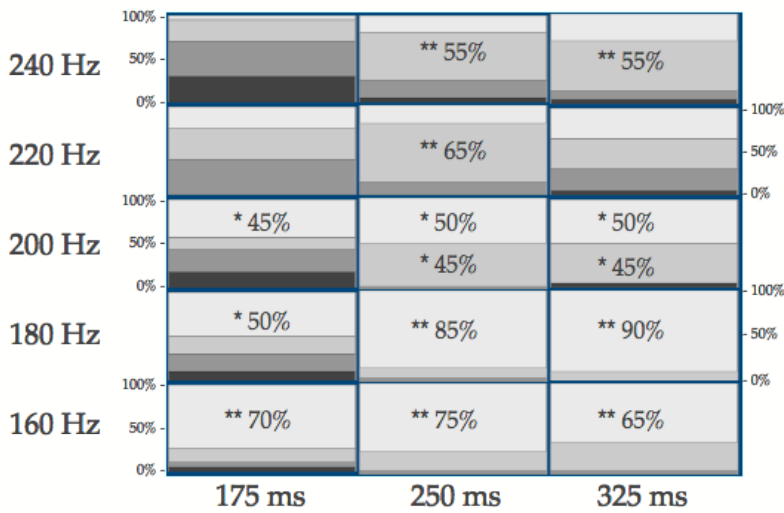
### 3. Results

Correct identification of tone for the unaltered recordings was high (160/180, 89%), with confusion between Creaky and Checked syllables yielding 11 of the 20 total misidentifications. The six prototype tokens were less successfully identified by phonation type. The modal prototype syllables were identified 100% as Low-toned, the creaky prototype 75% as Creaky-toned, but identifications of the breathy voiced prototype were split evenly between the Low (10/20) and High tone (10/20).

Identification of the synthesized tokens relied predominantly on F0 and duration for the breathy and modal sets of tokens. Creaky phonation stimuli were predominately identified as Creaky tone at all levels of F0 height or duration (cf Fig. 3). Results for the even contoured stimuli are presented graphically in Figures 1–3 below.

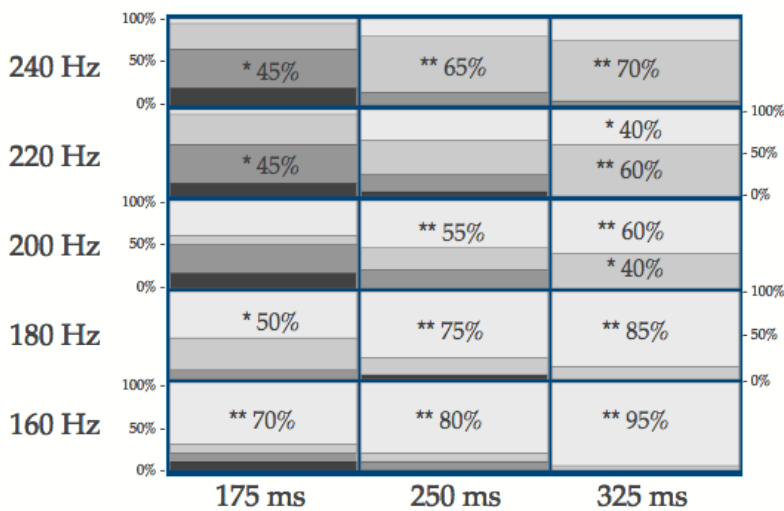
Above 200 Hz, breathy and modal stimuli were most often identified as High tones, while below 200 Hz Low tone identifications predominated. Precisely at 200 Hz, even contours were evenly split between the High and Low tone. At the shortest duration, 175 ms, breathy and modal stimuli at or over 200 Hz were also frequently identified as Creaky or Checked tone syllables.

Rising contours, naturally found just with High tones, produced similar results. Breathly and modal stimuli were most commonly identified as Low or High tones, and

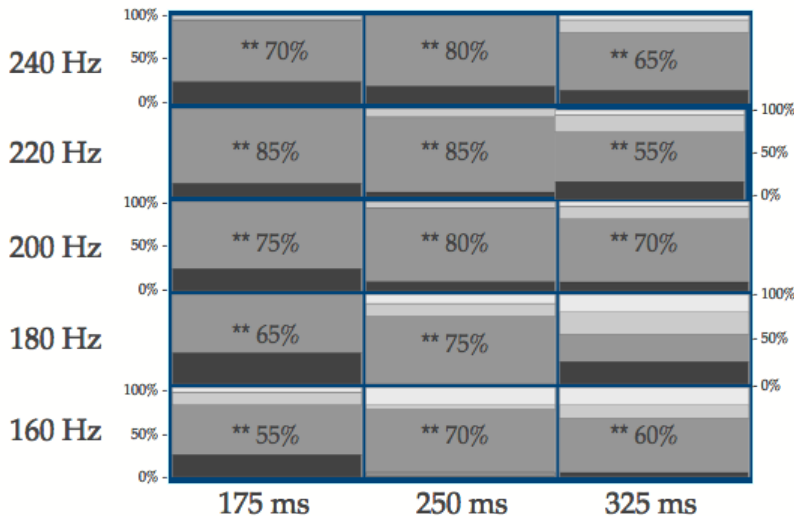


**Figure 1.**  
Breathy Stimuli  
Even F0 contours  
20 Hz intervals

In Fig. 1–3, each box represents 20 IDs. Rates of identification higher than chance are listed over the corresponding shaded portion (\*  $p < .05$ , \*\*  $p < .01$  by  $\chi^2$ ).



**Figure 2.**  
Modal Stimuli  
Even F0 contours  
20 Hz intervals



**Figure 3.**  
Creaky Stimuli  
Even F0 contours  
20 Hz intervals

creaky stimuli as Creaky tones at all durations. Identification of falling contours revealed a different pattern. Low tone identifications were rare. Breathy and modal stimuli again yielded results similar to one another – in this case divided between High and Creaky tone identifications dependent on duration. Falling creaky stimuli were overwhelmingly identified as Creaky tones.

A stepwise multinomial logistic regression was run with Duration, onset F0, mid-point F0, offset F0,  $\Delta$  F0, Phonation Type, Subject, and Vowel as factors. The first step found Phonation Type to be the factor that explained the greatest amount of variation in the response data (by itself correctly predicted 46.9% of all responses). Further iterations of the regression showed the next best predictors to be, in order, mid-point F0, Subject, Duration, Vowel Quality, and then  $\Delta$  F0.

#### 4. Summary

While these results indicate the importance of phonation type as an acoustic cue, notably only the distinction between creaky and non-creaky phonation influenced perception consistently. Listeners did not differentiate breathy stimuli from those with modal voicing. These results suggest that a single feature fails to define not only the four-way contrast, but also any individual category within that system. Rather than a single, primary cue for a tone, the evidence for Burmese implies that listeners utilize multiple aspects of the acoustic signal, such that the most reliable cue is likely context dependent. Additionally, the perception data reinforce a finding of some production accounts (Watkins 2005) that the tones' non-pitch behavior falls into two subsets: (1) the longer, non-falling High and Low tones, and (2) the shorter, falling, laryngealized Creaky and Checked tones. This binary opposition is argued to be more plausible from a perceptual standpoint than the three-way phonation opposition (breathy~modal~creaky) presumed in other accounts (Bradley 1982, Green 2005), and is also more in-line with two-way register contrasts frequently reported in languages of Southeast Asia (Gregerson 1976).

#### 5. References

- Bradley, David. 1982. Register in Burmese. In D. Bradley (Ed.) *Tonation* (Pacific Linguistics Series A-62), 117-132. Canberra: Australia National University.
- Forster, K.I. and J.C. Forster. 2009. DMDX: A Windows display program with millisecond accuracy. Monash University and University of Arizona.  
<http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm>
- Gregerson, Kenneth. 1976. Tongue root and register in Mon-Khmer. In P. Jenner, L. Thompson, & S. Starosta (Eds.) *Austroasiatic Studies I*. Oceanic Linguistics special publication 13: 323-70. Honolulu, Hawaii.
- Gruber, James. 2010. Understanding the role of voice quality in Burmese lexical tone: An aerodynamic and EGG study. Poster presentation at LabPhon 12. Albuquerque, NM.
- Javkin, Hector and Ian Maddieson. 1983. An inverse filtering analysis of Burmese creaky voice. In *UCLA Working Papers in Phonetics* 57: 115-125. Los Angeles: UCLA.
- Thein Tun. 1982. Some acoustic properties of tone in Burmese. In D. Bradley (Ed.) *Tonation* (Pacific Linguistics Series A-62), 117-132. Canberra: Australia National University.
- Watkins, Justin. 2001. Illustrations of the IPA: Burmese. *Journal of the International Phonetic Association* 31.2: 291-295.
- Watkins, Justin. 2005. Slippery Tones: some data on Burmese tones in context. Handout. Talk given at workshop *Making Sense of Prosody* at University College London.