2012. Proceedings of the Annual Meeting of the Berkeley Linguistics Society 38. DOI: http://dx.doi.org/10.3765/bls.v38i0.3328
Published for BLS by the Linguistic Society of America

Relative importance of phonation cues in White Hmong tone perception *

Marc Garellek, Patricia Keating, and Christina M. Esposito

UCLA Linguistics and Macalester Linguistics

1. Introduction

The study investigates the importance of phonation cues in White Hmong (henceforth, Hmong) tone identification. The tonal inventory of Hmong can be seen in Table 0.1, based on Esposito (2012). Hmong possesses seven productive tones, two of which involve non-modal phonation. The breathy (*-g*) tone is usually produced with a mid- or high-falling pitch contour similar to the high-falling modal (*-j*) tone (Esposito 2012). Therefore, the words *pog* [pɔ̤˩˥] 'grandmother' and *poj* [pɔ˩˥] 'female' differ mostly in phonation. The low modal (*-s*) tone and the low-falling (checked) creaky (*-m*) tones are characterized by phonation, pitch, and duration differences. Production studies have shown that the phonation differences between these two tones are large (Esposito 2012, Garellek 2012), although for at least some speakers, the low-falling (*-m*) tone is sometimes realized simply as a short, checked tone with *modal* phonation (Huffman 1987, Ratliff 1992). There is an additional non-productive tone, the *-d* tone, which is a syntactic variant of the *-m* tone.

Despite recent work on the production of non-modal phonation in certain Hmong tones, it is still unclear to what extent listeners use or rely on such changes in voice

*We would like to thank Susan Yang and members of the Hmong-American Partnership in St. Paul, MN, for their assistance in recruiting and testing participants. This work is part of a larger study conducted in collaboration with Jody Kreiman (UCLA Department of Head and Neck Surgery), and was supported by NSF grants BCS-0720304 and IIS-1018863, and NIH/NIDCD grant DC01797.

Table 0.1: Overview of White Hmong tones, from Esposito (2012).

Tone	Orthographic tone symbol	Example (IPA)	Example in Hmong orthography
High-rising	-b	[pɔ˧˥]	*pob* 'ball'
Mid	⊘	[pɔ˧]	*po* 'spleen'
Low	-s	[pɔ˩]	*pos* 'thorn'
High-falling	-j	[pɔ˥˩]	*poj* 'female'
Mid-rising	-v	[pɔ˧˥]	*pov* 'to throw'
Low-falling creaky	-m	[pɔ̰˩]	*pom* 'to see'
Mid- to high-falling breathy	-g	[pɔ̤˥˩]	*pog* 'grandmother'

quality. In a previous study on White Hmong and Green Mong perception, Andruski (2006) found that listeners were better at identifying natural tokens of the breathy and creaky tones than the low modal one, suggesting that they use phonation cues in tonal recognition. This hypothesis is consistent with findings on other tonal languages like Cantonese, Karen, Mandarin, and Vietnamese, where non-modal phonation (e.g. breathiness or glottalization/creakiness) accompanies certain tones (Belotel-Grenié and Grenié 1997, Brunelle 2009, Brunelle and Finkeldey 2011, Yu and Lam 2011). However, the relative importance of phonation compared to pitch and duration in Hmong tone identification is unknown.

To understand the relative role of phonation in White Hmong tone perception, we conducted a perceptual experiment where we manipulated F0 and duration, while keeping constant the original cues to breathy and creaky phonation (aside from those involving F0 changes). Given that the low-falling creaky tone in Hmong differs from the low modal tone not just in terms of phonation, but also in pitch and duration, we hypothesize that voice quality will be used in addition to duration and pitch cues, but its relative importance is unclear. In contrast, voice quality should be the major cue used for distinguishing the high-falling breathy vs. modal tones.

2. Method

2.1. Stimuli

Stimuli were produced from natural tokens of /pɔ/ with six of the seven possible tones, recorded in isolation by a female native speaker of Hmong. The string /pɔ/ in Hmong can form a licit word with any of the seven productive tones, as seen in Table 0.1. A summary of the F0 and duration manipulations used to create the stimuli is shown in Table 0.2, and further details can be found in Garellek *et al.* (under review).

Table 0.2: Summary of F0 manipulations used to create the stimuli.

Original token	Manipulation 1	Manipulation 2	Manipulation 3
[pɔ̤˨]	Flat F0 at different levels	F0 fall shortened	Entire contour lowered
[pɔ̰˩]	Kept short vs. lengthened	Lowered F0 of modal portion (for short and long stimuli)	Raised F0 of creaky portion (for short and long stimuli)
[pɔ˩]	Kept long vs. shortened	Raised F0 (for short and long stimuli))	Created F0 fall at end of stimulus (for short and long stimuli)
[pɔ˥] [pɔ˨] [pɔ˧]	Lowered F0	Raised F0	

The original breathy-toned stimulus (with a high-falling pitch contour) underwent three independent sets of pitch manipulations in order to obtain breathy tokens with varying F0 levels and contours. For the first set of manipulations, F0 was flattened to its starting high value (267 Hz) and then lowered successively in steps of 10 Hz steps to a minimum of 187 Hz. For the next set of manipulations, the starting high F0 of the original falling contour was lowered in steps of 10 Hz while keeping the end pitch constant, effectively decreasing the pitch change of the stimulus. For the stimulus with the lowest starting F0, the pitch change from start to end was only 10 Hz, compared with a fall of 60 Hz for the original breathy token. For the third set of manipulations, the entire original contour was lowered by 10 Hz increments, such that the final contour was low-falling instead of high-falling. However, in this set the pitch change in Hz from start to end did not differ across stimuli. In total, 25 stimuli were created from the original breathy-toned stimulus.

F0 manipulations were accomplished using the "Pitch-Synchronous Overlap and Add" (PSOLA) function in Praat, which alters F0 while preserving other spectral properties that can affect voice quality (Moulines and Charpentier 1990). This is done by separating the signal into discrete, overlapping segments, which are then repeated or omitted (for greater or lower F0, respectively). The remaining segments are finally overlapped and added together to reconstitute the speech signal.

The original modal and creaky words were first blocked according to length. Typically, the low modal tone is longer than the low creaky one, so a short version of the low-modal and a long version of the low-creaky words were created. Length of the vowel was manipulated in Praat by duplicating pulses from the middle of the vowel, which for both tones was modal-sounding. Low modal and low-falling

creaky stimuli with both original and modified durations then underwent two independent types of pitch modifications. For the low modal words, we first shifted the entire contour by 10 Hz increments between 120 and 210 Hz. In the other manipulation, we lowered the F0 of the original low-modal words to simulate the pitch fall of the low-falling creaky tone. At about two-thirds of the vowel's duration (which is when F0 typically begins to fall for the creaky tone), the pitch fell in 10 Hz increments to a maximum 70 Hz drop. The slope of the fall was created using quadratic interpolation in Praat, such that it dropped gradually. In total, 24 stimuli (12 long and 12 short) were created from the original low-modal stimulus.

We also performed two independent sets of F0 manipulations on the original creaky-toned word. In the first set of manipulations, we varied the pitch of the original creaky-toned stimuli by lowering the F0 of the non-creaky initial part of the vowel by increments of 10 Hz. In the second set of manipulations, we raised the F0 of the original creaky stimuli during the creaky portion (in the final third of the vowel) by 10 Hz increments, until the pitch was nearly flat. In total, 30 stimuli (15 long and 15 short) were created from the original creaky-toned word.

The other modal tones also underwent pitch manipulations. The whole F0 contour of the high and high-falling modal tones was lowered by 100 Hz in 10 Hz increments, and the F0 contour of the rising tone was raised up to 80 Hz in 20 Hz increments. In total, 38 stimuli were produced from the other modal tones: eight from the high-level tone, 20 from the high-falling modal tone, and 10 from the rising tone. The task had a total of 127 stimuli, each presented twice for a total of 254 tokens. Stimuli were randomized prior to each testing session.

An acoustic analysis for voice quality measures showed that, despite the F0 manipulations, the acoustic cues to the voice quality of the original sound had not been altered (cf. Esposito (2010)). H1*-H2*, H1*-A1*, and harmonics-to-noise ratio below 500 Hz (HNR) were used to analyze the tokens' voice quality, because these measures have been shown to distinguish modal phonation from both breathy and creaky phonation types in Hmong (Garellek 2012). H1*-H2* is the difference in amplitude of the first two harmonics, and H1*-A1* is the difference in the amplitude of the first harmonic and the harmonic nearest the first formant. The asterisks indicate that the measures have been corrected for the effects of vowel formants (Hanson 1995, Iseli *et al.* 2007). Breathy vowels are expected to have higher H1*-H2* and H1*-A1*, but lower values for HNR, than modal vowels. Creaky vowels are expected to have lower values than modal vowels for all three measures. We obtained these measures using VoiceSauce (Shue *et al.* 2011). As shown in Table 0.3, this is true for all stimuli, regardless of the F0 and duration manipulations. Thus, the phonation of the manipulated stimuli were characteristic of breathy, modal, and creaky voice quality in Hmong.

Table 0.3: Mean values of H1*-H2*, H1*-A1*, and HNR in dB (standard deviations in parentheses) for high-falling breathy vs. modal and low creaky vs. low modal stimuli, across all pitch manipulations.

	H1*-H2*	H1*-A1*	HNR
High-falling breathy	8.36 (3.37)	27.11 (5.23)	27.67 (1.06)
High-falling modal	3.83 (2.18)	22.15 (1.34)	38.08 (4.39)
Low-falling creaky	1.40 (1.30)	21.68 (1.05)	35.94 (6.09)
Low modal	5.03 (1.96)	28.56 (2.97)	37.48 (6.12)

2.2. Participants

Participants were recruited at the Hmong-American Partnership and through personal contacts in St. Paul, Minnesota. Fifteen native speakers of White Hmong, eight men and seven women, participated in the experiment. All spoke English with varying degrees of proficiency, and all spoke Hmong daily, both at work and at home. They were all literate in Hmong Romanized Popular Alphabet (R. P. A.) script. The experiment lasted about 20-30 minutes and was conducted in a quiet room. Participants listened to the experiment using noise-attenuating headphones. They were compensated for their time.

2.3. Task

The experiment was implemented in Praat (Boersma and Weenink 2011), and consisted of a seven-alternative forced-choice identification task, during which participants listened to stimuli, and then indicated which word of the form /pɔ/ they heard. The possible words were displayed on screen in standard Hmong orthography, which uses letters after the vowel to mark the tone, except for the mid tone, which is not marked orthographically. Listeners could hear the stimulus as many times as they wished before selecting their response, which they were able to change before hearing the next stimulus. A bilingual English-Hmong experimenter ensured that the participants understood the task.

3. Results

Participants' responses were analyzed using logistic mixed-effects regression to determine the relevant factors that account for choosing a breathy or creaky response. The regression was done in R using the *lmer* function in the *lme4* package (Baayen

2008). The original phonation of the word was coded as being either breathy, modal, or creaky, according to the lexical tone.

For predicting 'breathy tone' responses, the logistic model included the original phonation of the stimulus (breathy vs. non-breathy), the F0 averaged over the first ninth of the vowel, the F0 averaged over the final ninth, whether the F0 was flat vs. a contour, and mean F0. The F0 was measured in the first and final ninths of the vowel in order to get start and end values of the measure. Average F0 values over short intervals were used (instead of values at single time points) in order to smooth the data. Participant was included as a random effect, and the dependent variable was whether or not participants chose a 'breathy tone' response. The results are shown in Table 0.4. Of the fixed effects, the only significant factor was whether the original stimulus was breathy, which significantly increased the likelihood of a 'breathy tone' response ($p<0.0001$).

Table 0.4: Fixed-effects results of logistic model predicting 'breathy tone' responses.

	Estimate	SE	*Z*-score	*p*-value
Intercept	-2.48	0.38	-6.58	<0.0001***
Orig. tone=breathy	3.98	0.18	21.57	<0.0001***
Mean F0	-0.0008	0.01	-0.09	0.93
F0 in 1st ninth	-0.01	0.01	-1.91	0.06
F0 in final ninth	0.01	0.01	1.02	0.31
F0 slope - flat	-0.04	0.16	-0.27	0.79

For predicting 'creaky tone' responses, the logistic model included the original phonation of the stimulus (creaky vs. non-creaky), the stimulus length (short vs. long), the F0 averaged over the first ninth, the F0 during the final ninth, slope of F0 (contour vs. flat), and mean F0. Participant was included as a random effect, and the dependent variable was whether or not participants chose a 'creaky tone' response. The results are shown in Table 0.5. The phonation of the original stimulus did not matter, even if it was creaky. Instead, the F0 in the final ninth, the F0 slope, and the stimulus length were significant (all $p<0.001$). Thus, a stimulus that was short in length, with a non-flat F0 contour, and/or a lower final F0 was associated with overall greater 'creaky tone' responses.

4. Discussion

The issue of whether non-modal phonation plays a primary role in tone perception in languages with and without phonation contrasts is understudied. In the case

Table 0.5: Fixed-effects results of logistic model predicting 'creaky tone' responses.

	Estimate	SE	*Z*-score	*p*-value
Intercept	1.30	0.37	3.55	<0.001**
Orig. tone=creaky	0.09	0.14	0.61	0.54
Mean F0	-0.005	0.01	-0.94	0.35
F0 in 1st ninth	-0.001	0.004	-0.34	0.73
F0 in final ninth	-0.02	0.004	-3.45	<0.001**
F0 slope - flat	-1.10	0.18	-6.18	<0.0001***
Length - short	1.11	0.13	8.69	<0.0001***

of White Hmong, phonation cues are fundamental for identifying the high-falling breathy tone, with F0 modifications having little effect. On the other hand, the role of phonation in the identification of the low-falling creaky tone is apparently minor, given that an F0 dip and short duration are what listeners relied on in this study. Therefore, whereas breathiness is used to make a categorical distinction between two tones, creakiness likely reinforces the F0 lowering and short duration of the low-falling creaky/checked tone. In this way, creaky voice quality appears to be a secondary cue to what is fundamentally a duration and pitch contrast. This might seem surprising, given the evidence that creakiness in Hmong is extensive in both time and degree (Esposito 2012, Garellek 2012).

The results show that participants treated breathiness and creakiness differently. Breathiness was independent of F0, such that pitch modulations of breathy stimuli did not change participants' responses. Thus, participants still perceived a flat F0 (at various pitch heights) as breathy, even though in natural speech the breathy tone in Hmong is produced with a falling pitch contour. If a breathy-toned vowel was low-falling instead of the more natural high-falling pitch contour, participants still perceived it as breathy, as shown in Figure 0.1. We found no significant change in 'breathy tone' responses when the starting F0 varied, even when its pitch contour resembled that of the creaky tone more than the modal or breathy high-falling tones.

On the other hand, identification of the creaky tone in Hmong was highly dependent on the duration and F0 of the stimulus. For participants to identify a word as creaky-toned, the vowel needed to be short and have a low-falling pitch contour, but creaky voice quality (aperiodic and with low spectral tilt) was not necessary. This is demonstrated in Figure 0.2, which plots proportion 'creaky' responses as a function of the pitch fall for short original creaky and low modal stimuli. There was little difference between the original creaky and low modal tokens with manipulated

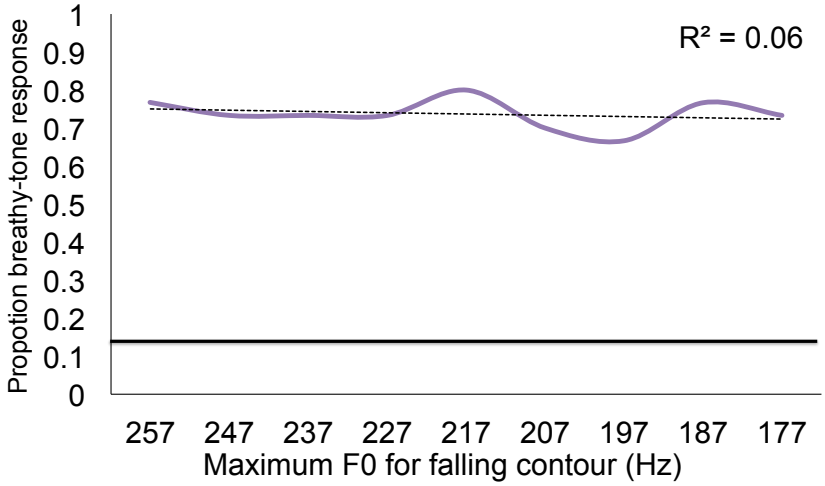
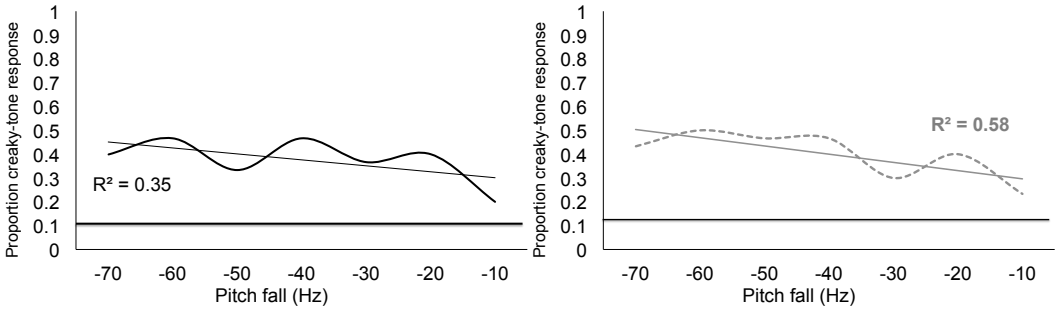


Figure 0.1: Proportion 'breathy' responses for breathy stimuli as a function of start of F0 fall. Chance is indicated at 0.14.

F0, with both groups identified as creaky only about 40% of the time. For both categories there was a moderate correlation between 'creaky tone' responses and the pitch fall, consistent with the logistic regression results. The absence of a difference between the modal and creaky stimuli shows that presence of creaky phonation in the original token mattered little in the prediction of 'creaky' responses. Note also that the overall creaky-tone identification rate was lower than for the breathy tone. This suggests that the stimuli were not ideal creaky-tone tokens, or that the low-falling creaky tone is in general more confusable with other tones.



(a) Short creaky-tone stimuli

(b) Shortened low-modal stimuli

Figure 0.2: Proportion 'creaky' responses as a function of start of F0 fall, for low-falling creaky (*-m*) and shortened original low modal (*-s*) tokens. Chance is indicated at 0.14.

However, we do not claim that the low-falling tone in Hmong need not be creaky at all. Rapid dips in F0 can cue creaky voice quality in Mixtec (Gerfen and Baker

2005) and glottal stops in English (Hillenbrand and Houde 1996), suggesting that some forms of creaky voice can be tied to pitch dynamics alone. Therefore, this study reinforces this fundamental distinction between breathy voice, which is pitch-independent, and creaky voice, which in some forms is a type of pitch setting.

5. Conlusions

This study shows that Hmong listeners used breathy voice quality when differentiating the breathy and modal high-falling tones. Creaky voice quality, however, is not a necessary cue for the low creaky tone, perhaps because the pitch fall and shorter duration are sufficient for listeners to identify it as distinct from the low modal tone. Creaky phonation can therefore be seen as a means of reinforcing the low F0 target at the end of the low-falling tone (and maybe also its checked-like short duration). In this way, it seems that in Hmong 'breathy' is contrastive in a way that 'creaky' is not.

References

Andruski, Jean E., 2006. *Tone clarity in mixed pitch/phonation-type tones*. *Journal of Phonetics* 34:388–404.

Baayen, R. H., 2008. *Analyzing Linguistic Data. A practical introduction to statistics.* Cambridge: Cambridge University Press.

Belotel-Grenié, Agnès and Grenié, Michel, 1997. *Types de phonation et tons en chinois standard*. *Cahiers de linguistique - Asie orientale* 26:249–279.

Boersma, Paul and Weenink, David, 2011. *Praat: doing phonetics by computer [Computer program]. Version 5.3.02.* Retrieved November 10, 2011 from http://www.praat.org/.

Brunelle, Marc, 2009. *Tone perception in Northern and Southern Vietnamese*. *Journal of Phonetics* 37:79–96.

Brunelle, Marc and Finkeldey, Joshua, 2011. *Tone perception in Sgaw Karen*. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 17)*, 372–375.

Esposito, Christina M., 2010. *The effects of linguistic experience on the perception of phonation*. *Journal of Phonetics* 38:306–316.

Esposito, Christina M., 2012. *An acoustic and electroglottographic study of White Hmong phonation*. *Journal of Phonetics* 40:466–476.

Garellek, Marc, 2012. *The timing and sequencing of coarticulated non-modal phonation in English and White Hmong*. *Journal of Phonetics* 40:152–161.

Garellek, Marc, Esposito, Christina M., Keating, Patricia, and Kreiman, Jody, under review. *Perception of spectral slopes and White Hmong tone identification*.

Gerfen, Chip and Baker, Kirk, 2005. *The production and perception of laryngealized vowels in Coatzospan Mixtec*. *Journal of Phonetics* 33:311–334.

Hanson, Helen M., 1995. *Glottal characteristics of female speakers*. Ph.D. thesis, Harvard University.

Hillenbrand, James M. and Houde, Robert A., 1996. *Role of F0 and amplitude in the perception of glottal stops*. *Journal of Speech and Hearing Research* 39:1182–1190.

Huffman, Marie K., 1987. *Measures of phonation type in Hmong*. *Journal of the Acoustical Society of America* 81:495–504.

Iseli, Markus, Shue, Yen-Liang, and Alwan, Abeer, 2007. *Age, sex, and vowel dependencies of acoustic measures related to the voice source*. *Journal of the Acoustical Society of America* 121:2283–2295.

Moulines, Eric and Charpentier, Francis, 1990. *Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones*. *Speech Communication* 9:453–467.

Ratliff, Martha, 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.

Shue, Yen-Liang, Keating, Patricia A., Vicenik, Chad, and Yu, Kristine, 2011. *VoiceSauce: A program for voice analysis*. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS 17)*, 1846–1849. Hong Kong.

Yu, Kristine M. and Lam, Hiu Wai, 2011. *The role of creaky voice in Cantonese tonal perception*. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 17)*, 2240–2243.

Marc Garellek
Phonetics Laboratory, Department of Linguistics
University of California, Los Angeles
3125 Campbell Hall, Los Angeles, CA 90095-1543

marcgarellek@ucla.edu