Voice quality and tone identification in White Hmong

Marc Garellek[a] and Patricia Keating
Phonetics Laboratory, Department of Linguistics, University of California, Los Angeles, California, 90095-1543

Christina M. Esposito
Department of Linguistics, Macalester College, St. Paul, Minnesota, 55105

Jody Kreiman
Department of Head and Neck Surgery, School of Medicine, University of California, Los Angeles, California, 90095-1794

(Received 9 February 2012; revised 18 November 2012; accepted 29 November 2012)

This study investigates the importance of source spectrum slopes in the perception of phonation by White Hmong listeners. In White Hmong, nonmodal phonation (breathy or creaky voice) accompanies certain lexical tones, but its importance in tonal contrasts is unclear. In this study, native listeners participated in two perceptual tasks, in which they were asked to identify the word they heard. In the first task, participants heard natural stimuli with manipulated F0 and duration (phonation unchanged). Results indicate that phonation is important in identifying the breathy tone, but not the creaky tone. Thus, breathiness can be viewed as contrastive in White Hmong. Next, to understand which parts of the source spectrum listeners use to perceive contrastive breathy phonation, source spectrum slopes were manipulated in the second task to create stimuli ranging from modal to breathy sounding, with F0 held constant. Results indicate that changes in H1-H2 (difference in amplitude between the first and second harmonics) and H2-H4 (difference in amplitude between the second and fourth harmonics) are independently important for distinguishing breathy from modal phonation, consistent with the view that the percept of breathiness is influenced by a steep drop in harmonic energy in the lower frequencies.
[http://dx.doi.org/10.1121/1.4773259]

PACS number(s): 43.71.Hw, 43.71.Es, 43.71.Bp, 43.70.Mn [BRM] Pages: 1078–1089

I. INTRODUCTION

Individual speakers' voices differ, and these differences can provide useful information for listeners. However, it is currently unclear which aspects of voice matter to listeners. Kreiman *et al*. (2007) analyzed a set of 70 normal and pathological voices, comparing acoustic measures of their source spectra by principal components analysis. In one analysis, slopes of portions of the spectral envelope were compared; in another, amplitude differences between various pairs of harmonics, and the amplitude of high-frequency noise, were compared. In both analyses, the lowest-frequency region (corresponding to H1-H2, the difference in amplitudes of the first and second harmonics) and the highest-frequency region (corresponding to high-frequency noise), were independently important in accounting for variance in the voices. Further independent and significant components in the two analyses corresponded to slopes or differences over a variety of smaller portions of the spectrum.

This statistical model of Kreiman *et al*. (2007) now needs to be developed into a perceptually valid model of listeners. The statistical model's many significant components are unlikely to all be perceptually relevant to listeners, and therefore, some initial simplifications are in order. Here, the spectrum above the second harmonic is divided into three parameters. One is H2–H4 (the difference in amplitudes of the second and fourth harmonics), which emerged as a significant component in the analysis of harmonic amplitude differences. Then the spectrum from H4 to 5 kHz is simply divided into two larger parameters: from H4 to the harmonic nearest 2 kHz (H4–2 kHz), and from 2 kHz to the harmonic nearest 5 kHz (2 kHz–5 kHz). The four resulting harmonic-amplitude parameters (H1–H2, H2–H4, H4–2 kHz, 2 kHz–5 kHz) cover the entire frequency range to 5 kHz. When spectral noise is added to the harmonic-amplitude parameters, a five-parameter model of the voice source spectrum is obtained. Here we focus on the perception of just the harmonic amplitudes, leaving aside the important question of how they interact perceptually with spectral noise in the perception of voice quality [though see Kreiman and Gerratt (2005), Shrivastav and Sapienza (2006), and Kreiman and Gerratt (2012) for initial results on that topic]. This simple model might suffice to describe cross-speaker voice differences.

In some languages, however, voice quality is not only a matter of individual voice differences; in addition, it defines linguistic contrasts. In Jalapa Mazatec, for example, each vowel in a word is specified as having modal, breathy, or creaky (laryngealized) phonation (Silverman *et al*., 1995; Garellek and Keating, 2011). In White Hmong, words can differ in creaky vs modal phonation, or breathy vs modal phonation. Because languages which employ nonmodal

[a]Author to whom correspondence should be addressed. Electronic mail: marcgarellek@ucla.edu

 0001-4966/2013/133(2)/1078/12/$30.00

phonations contrastively may use any of a variety of acoustic parameters to do so (Gordon and Ladefoged, 2001; Keating *et al.*, 2011), it is possible that a source spectrum model derived from individual voice differences will not extend to the description of linguistic contrasts. To further the development of the model along perceptual lines, then, it is important to apply it to the perception of phonation contrasts in languages like White Hmong. One goal of the current study is to test whether the harmonic-amplitude parameters of the model suffice to account for the perception of these contrasts.

However, the matter is complicated by the fact that, in some (often so-called "register") languages, phonation differences are tied to pitch differences, so that a lexical "tone" can be a combination of pitch plus nonmodal phonation. For example, in Santa Ana del Valle Zapotec, high- and rising-pitch tones have modal phonation, while falling-pitch tones have creaky or breathy phonation (Esposito, 2010b). In Northern Vietnamese, certain tones are reported to have some laryngealization or breathiness (Pham, 2003; Michaud, 2004; Brunelle, 2009). In Mandarin, Tone 3 has a low (or falling) pitch contour, and often also creaky voice (Belotel-Grenié and Grenié, 1997). In such cases, it is not always clear whether listeners pay attention only to phonation, only to pitch, or to phonation as well as pitch. Studies of a few languages have begun to show that sometimes listeners do attend to phonation, even when it is correlated with pitch (Belotel-Grenié and Grenié, 1997; Brunelle, 2009; Brunelle and Finkeldey, 2011; Yu and Lam, 2011; Brunelle, 2012). But in other cases, listeners do not attend to phonation (notably breathiness), preferring pitch information exclusively (Brunelle, 2009; Brunelle and Finkeldey, 2011; Brunelle, 2012).

The Hmongic languages are similar to those just mentioned, in that they have some lexical tones that involve both phonation and pitch differences (Huffman, 1987; Ratliff, 1992; Andruski and Ratliff, 2000; Andruski, 2006; Fulop and Golston, 2008; Esposito, 2012; Garellek, 2012). The tonal inventory of White Hmong can be seen in Table I, based on Esposito (2012). In Hmong the tone is marked orthographically by a letter at the end of each syllable. The words [pɔ̰˩˧] "grandmother" (g-tone) and [pɔ˩˧] "female" (j-tone) have similar F0 contours and durations, and so they differ mostly in phonation. In contrast, the words [pɔ˧] "thorn" (s-tone) and [pɔ̰˧˩] "to see" (m-tone) are distinguished by phonation, pitch, and duration differences. Production studies have shown that the phonation differences between these two tones are robust (Esposito, 2012; Garellek, 2012), though sometimes limited in duration, with the low-falling (*-m*) tone sometimes realized as a partly modal vowel followed by some glottalization and a glottal stop (Huffman, 1987; Ratliff, 1992; Esposito, 2012). Strong creak is not consistently found for the low-falling (*-m*) tone, and the phonation difference between the two tones can be more like tense vs modal voice. Pitch and duration differences have also been shown, with the low-falling (*-m*) tone having a lower final pitch and a shorter duration than any other tones (Huffman, 1987; Esposito, 2012).

TABLE I. Overview of White Hmong tones, from Esposito (2012).

Tone	Orthographic tone symbol	Example (IPA)	Example in Hmong orthography with English meaning
High-rising (45)	-b	[pɔ˧˥]	*pob* "ball"
Mid (33)	∅	[pɔ˧]	*po* "spleen"
Low (22)	-s	[pɔ˨]	*pos* "thorn"
High-falling (52)	-j	[pɔ˥˩]	*poj* "female"
Mid-rising (24)	-v	[pɔ˧˥]	*pov* "to throw"
Low-falling creaky (21)	-m	[pɔ̰˨˩]	*pom* "to see"
Mid-to high-falling breathy (52 or 42)	-g	[pɔ̤˥˩]	*pog* "grandmother"

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, indicating that the breathy and creaky tones have reliable and salient acoustic correlates. Because her stimuli were unaltered natural tokens, listeners had access to all the naturally occurring acoustic information for each tone. As a result, the relative importance of phonation compared to pitch and duration information cannot be determined from that experiment. Another goal of the present study is to tease apart the perceptual contributions of these acoustic properties. We will thus gain a clearer understanding of the pitch-phonation interactions in the Hmong tone system, which in turn will contribute to our understanding of such tone systems more generally.

In sum, before testing which aspects of spectral slope underlie phonation perception in White Hmong, we want to be sure that White Hmong listeners are attending to phonation in the first place. Therefore, we must first demonstrate that, given a choice of attending to pitch vs to phonation, listeners in fact do the latter. In experiment 1, we pit pitch against phonation in the identification of the two contrasts in White Hmong illustrated above, and reveal that for one contrast, but not the other, listeners do use phonation information. In experiment 2, we then test the harmonic-amplitude parameters in the perception of this contrast.

II. EXPERIMENT 1: F0 AND DURATION MANIPULATIONS

Experiment 1 examines the role of original voice quality relative to F0 and duration manipulations in breathy and creaky tone identification. The two tone pairs of interest here are the contrast between the two high-falling tones (e.g., [pɔ˥˩] "female" vs [pɔ̤˥˩] "grandmother") and the contrast between the two low-pitch tones (e.g., [pɔ˧] "thorn" vs pɔ̰˧˩] "to see"). In this experiment, we performed various F0 manipulations to naturally creaky and breathy stimuli (leaving the phonation intact), and then determined if the nonmodal phonation suffices for perception of the breathy or creaky tone.

A. Method

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 White Hmong. A clear token of mid-level-toned /pɔ˧/ recorded in isolation was not obtained. However, through pitch resynthesis there were stimuli with

pitch contours typical of the mid-level tone. The string /pɔ/ in White Hmong can form a licit word with any of the seven tones, as seen in Table I.

In order to assess the role of breathy voice quality in the perception of the high-falling breathy tone, the original breathy-toned stimulus (with a high-falling pitch contour) underwent three independent sets of pitch manipulations which were meant to obtain breathy tokens with varying F0 levels and contours, many of which would be unlikely for naturally occurring breathy-toned vowels. These manipulations are schematized in Fig. 1. 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 to a minimum of 187 Hz. These manipulations were meant to render breathy vowels confusable with the level tones in Hmong, which are all modal. For the next set of manipulations, the final F0 of the original falling contour was raised in steps of 10 Hz while keeping the starting pitch constant, effectively decreasing the pitch change of the stimulus and thus possibly rendering it more confusable with other modal level tones. For the stimulus with the highest final 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, which would make the pitch contour more similar to that of the low-falling creaky tone. 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.

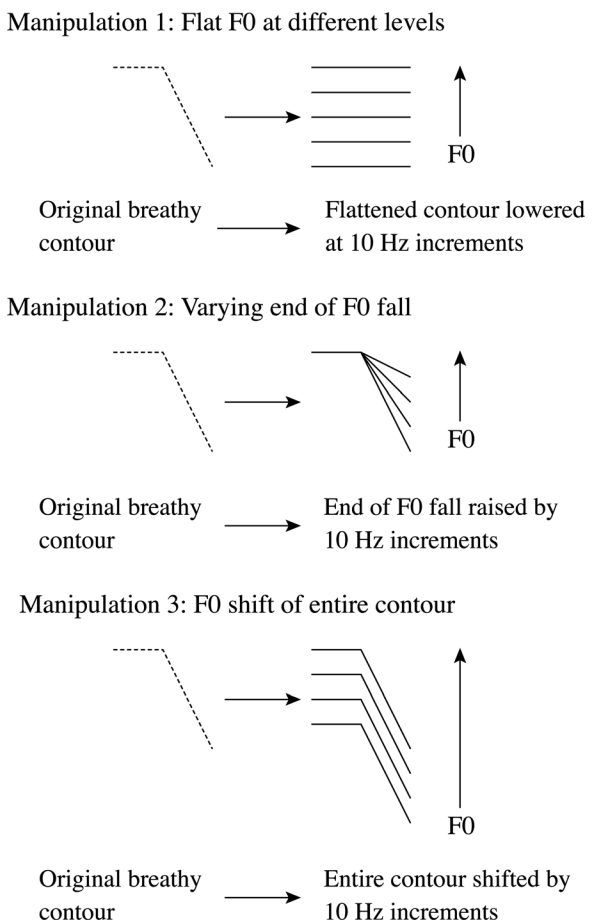


FIG. 1. Schematic F0 contours for the three sets of manipulations performed on the original breathy-toned stimuli. The upward facing arrow indicates the direction of F0.

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.

In order to assess the role of creaky voice quality in the perception of the low-falling creaky tone, we manipulated the duration and F0 of the original creaky stimulus. 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 (200 ms) of the low-modal and a long version (337 ms) of the low-creaky words were created to determine what role vowel length plays in identifying the creaky tone. 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. This was done in order to obtain a variety of low-pitched stimuli, some of which would be lower than expected for the low-modal tone, which might render it more confusable with the low-creaky tone. In the other manipulation, we lowered the final F0 of the original low-modal words to simulate the pitch fall of the low-falling creaky tone, potentially rendering these low modal stimuli more confusable with the 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. A schematic of the two sets of manipulations for original low-modal stimulus is shown in Fig. 2(a). 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 noncreaky initial part of the vowel by increments of 10 Hz, so that some creaky stimuli would have little to no pitch fall. 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, so that some stimuli were creaky but not very low in pitch. A schematic of the two sets of manipulations for original creaky-toned stimuli is shown in Fig. 2(b). 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, in order to provide a variety of words with modal phonation and varying F0 contours. The whole F0 contour of the high and high-falling modal tones was lowered by 100 Hz in

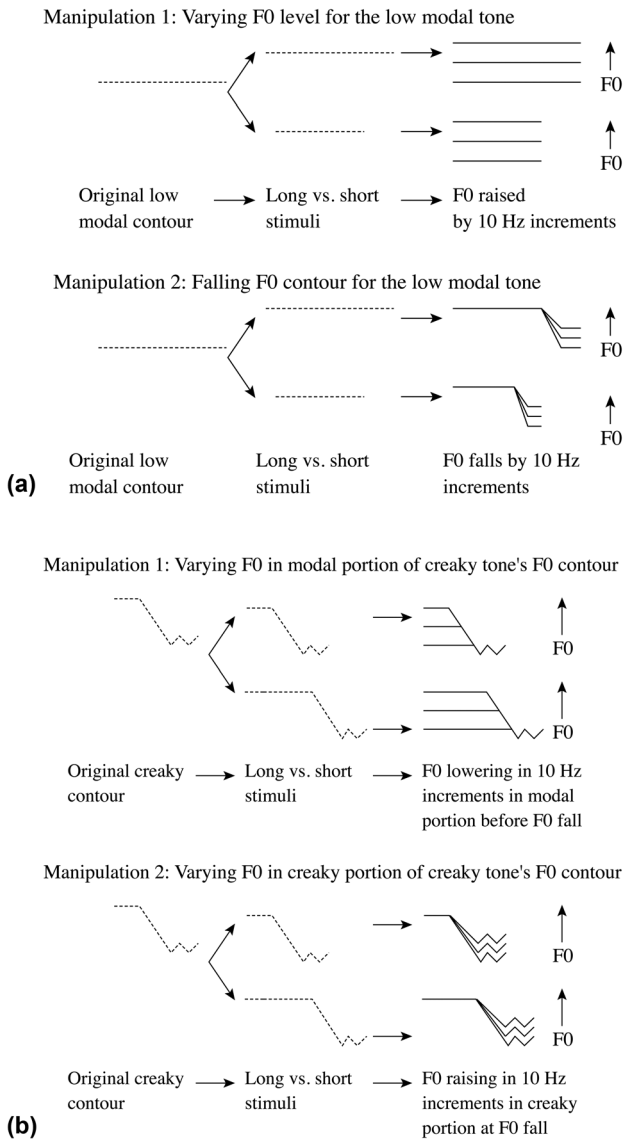


FIG. 2. Schematic F0 contours for the two sets of manipulations performed on (a) the original low modal-toned stimuli, and (b) the original creaky-toned stimuli. The upward-facing arrow indicates increasing F0, and the jagged lines represent creaky phonation.

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 thus included 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, 2010a). H1*–H2*, H1*–A1*,[1] 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 White Hmong (Garellek, 2012). 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

TABLE II. Mean values of H1*–H2*, H1*–A1* (asterisks indicate that the measures are corrected for formant frequencies and bandwidths), 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)

three measures. We obtained these measures using VoiceSauce (Shue *et al.*, 2011). As shown in Table II, the expected differences in the acoustic measures by phonation type hold for all stimuli, regardless of the F0 and duration manipulations. Thus, the phonation of the manipulated stimuli was characteristic of breathy, modal, and creaky voice quality in White Hmong.

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, aged 25–55, participated in the experiment. All spoke English, though with varying degrees of proficiency, and all spoke White Hmong daily, both at work and at home. They were born in Thailand, Laos, or the US, and all were literate in the Hmong Romanized Popular Alphabet (R.P.A.) script, which they used on a daily basis. None reported any difficulty with reading the words displayed or understanding the task. The experiment lasted about 20–30 min and was conducted in a quiet room. Participants listened to the stimuli over noise-attenuating headphones. They were compensated for their time.

3. Task

Experiment 1, which was implemented in Praat (Boersma and Weenink, 2011), consisted of a seven-alternative forced-choice identification task, during which participants listened to stimuli of form /pɔ/, and then indicated which word 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-White Hmong experimenter ensured that the participants understood the task.

B. Results

Participants' responses were analyzed using logistic mixed-effects regression to determine the relevant factors that account for choosing a breathy-toned or creaky-tone response. Thus, responses were re-coded as binomial variables, by combining all "non-breathy-toned" responses (to compare with "breathy-toned" responses), or by combining all "non-creaky-toned" responses (to compare with "creaky-toned" responses). The regression was done in R using the

lmer function in the *lme4* package with a binomial distribution and logistic link function selected for the generalized linear mixed-effect model. P-values are provided in the output of the generalized linear mixed-effects model used for logistic regression (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-toned" 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. These F0 factors were chosen because together they represented the crucial dimensions in which stimuli could vary in pitch: overall pitch, pitch dynamics (flat vs contour), and start and end values. Only random intercepts by participant were included, because no larger random structure significantly improved model fit, which was assessed by model comparison using anova. The dependent variable was whether or not participants chose a "breathy-toned" response.

The results of the logistic regression model are shown in Table III. Of the fixed effects, the only significant factor was whether the original stimulus was breathy, which significantly increased the likelihood of a "breathy-toned" response ($p < 0.0001$).[2] The effect of the F0 in the first ninth was close to significant (at $p = 0.06$), but minor compared to the original phonation of the stimulus, as indicated by the much larger *Z*-score of the latter compared to the former.

For predicting "creaky-toned" 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 averaged over 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-toned" response. The results are shown in Table IV. The phonation of the original stimulus did not matter, even if it was creaky. Instead, the effects of 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 low final F0 was associated with an overall greater number of "creaky-toned" responses.[3]

TABLE III. Fixed-effects results of logistic model predicting "breathy-toned" responses.

	Coefficient β	SE (β)	*Z*-score	*p*-value
Intercept	−2.48	0.38	−6.58	<0.0001
Presence of breathy phonation	3.98	0.18	21.57	<0.0001
Mean F0	−0.001	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

TABLE IV. Fixed-effects results of logistic model predicting "creaky-toned" responses.

	Coefficient β	SE (β)	*Z*-score	*p*-value
Intercept	1.30	0.37	3.55	<0.001
Presence of creakiness	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

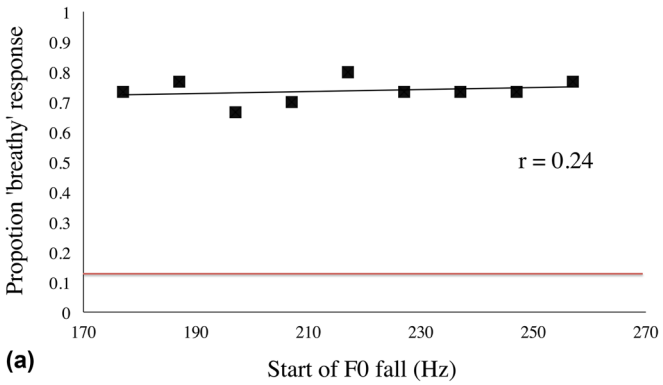
C. Discussion

The results from experiment 1 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 White 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-toned, as shown in Fig. 3(a). We found no significant change in "breathy-toned" 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 White 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 Fig. 3(b), which shows the proportion of "creaky-toned" responses as a function of the pitch fall for short original creaky and low modal stimuli. There are few differences between the original creaky and low modal tokens with manipulated F0, with both groups identified as creaky only about 40% of the time. For both categories there was a moderate correlation between "creaky-toned" responses and the pitch fall, consistent with the logistic regression results. The absence of a difference between the modal and creaky stimuli shows that the presence of creaky phonation in the original token mattered little in the prediction of "creaky-toned" responses.

III. EXPERIMENT 2: SOURCE SPECTRUM MANIPULATIONS

The results of experiment 1 show that only the breathy-modal contrast appears to be sufficiently contrastive to lend itself to testing the source spectrum model; the phonation component of the creaky-modal contrast does not sufficiently engage listeners' attention. Given these results, the next question of interest is to determine which parts of the source spectrum listeners use to perceive contrastive breathy phonation. Therefore, in experiment 2, source spectrum slopes were manipulated to create stimuli ranging from modal to breathy sounding, with F0 held constant.



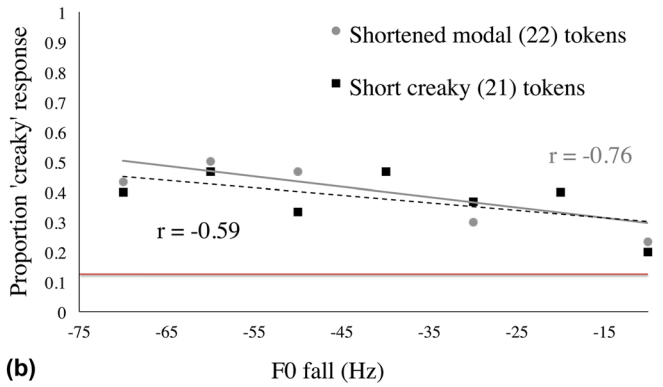


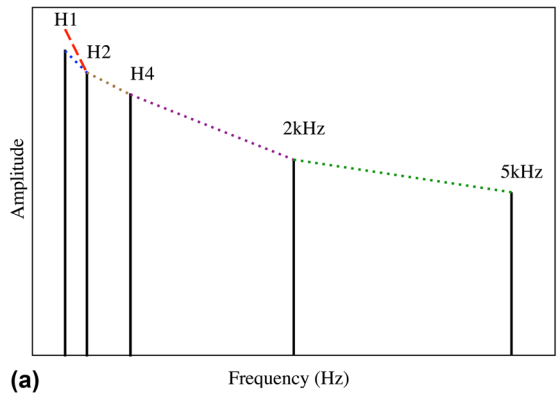
FIG. 3. (Color online) (a) Proportion "breathy-toned" responses for breathy stimuli as a function of start of F0 fall. (b) Proportion "creaky-toned" responses as a function of start of F0 fall, for short original low modal (22) and low-falling creaky (21) tokens. Chance is indicated at 0.14. Best-fit regression lines are shown for (a) proportion breathy response as a function of F0 fall (excluding subject-explained variance), and (b) for shortened modal (grey line) and short creaky (black line) stimuli. The starting value of the F0 fall is weakly correlated with proportion "breathy-toned" responses. For "creaky-toned" responses, F0 fall shows moderate to high correlations with proportion "creaky-toned" responses for *both* modal and creaky stimuli, as indicated by Pearson's *r* values. Note that these correlations are not derived from the logistic mixed-effects models.

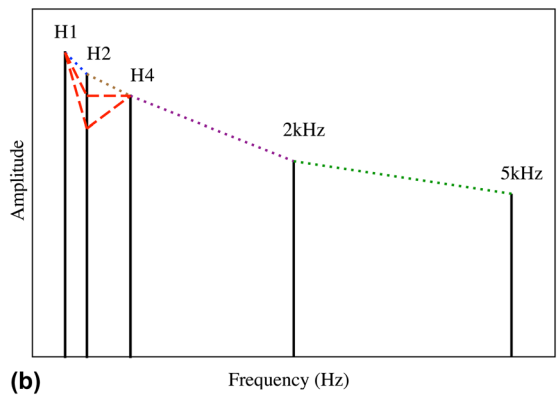
A. Method

1. Stimuli

The stimuli were produced using the UCLA voice synthesizer (Kreiman *et al.*, 2010; Kreiman and Gerratt, 2010). A 1-s sample of a sustained /a/ by a female English speaker (with an F0 of 230 Hz) was copied such that the synthesized vowel copy formed a good match to the original, in terms of both acoustic and perceptual characteristics. We used an English speaker's vowel because we did not have a recorded Hmong token suitable for inverse filtering and copy-synthesizing. However, the token's formants (F1 = 780 Hz, F2 = 1330 Hz) fell within the normal range for Hmong /a/ with either of the high-falling tones, based on the Hmong tokens in Esposito (2012). And, because the source is entirely manipulated for the experiment, the initial spectral profile mattered little. The synthesized token was shortened to a duration appropriate for both of the high-falling tones in White Hmong (about 340 ms), and the fundamental frequency was adjusted (by the method described in Sec. II A 1) such that it was high-falling from 280 to 198 Hz. This F0 contour was taken from a natural token of White Hmong *tag* /tą̈˥/ "finish" spoken by a White Hmong female speaker. All stimuli were produced from this synthesized base /ą̈˥/.

The source shape of this base form was then modified according to the four harmonic-amplitude parameters of the source model described in the Introduction (Kreiman *et al.*, 2007, 2011; Kreiman and Gerratt, 2012): H1–H2, H2–H4, H4–2 kHz (the harmonic closest to 2000 Hz), and 2 kHz–5 kHz (the final harmonic in the source model, closest to 5000 Hz).[4] For every stimulus file, the amplitudes of all harmonics were adjusted to the slopes of these components, as shown in Fig. 4. Only the harmonic amplitudes were modified; the noise component of the original sample was set at a





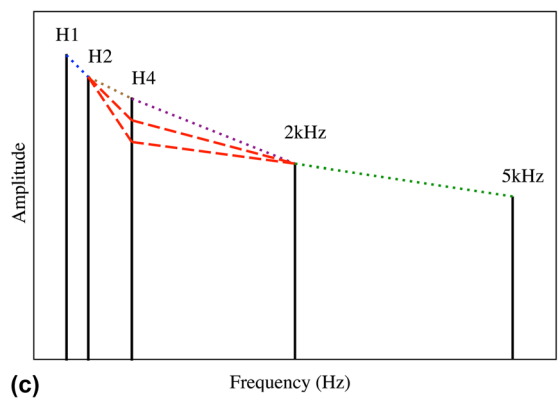


FIG. 4. (Color online) Schematics of the spectral source manipulations, with only the five harmonic anchors (H1, H2, H4, 2 kHz, and 5 kHz) represented by vertical lines. The four component slopes of the source spectrum model are shown by dotted lines. In condition 1 (a), H1–H2 varied by increasing the amplitude of H1. In condition 2 (b), H1–H2 and H2–H4 covaried by decreasing the amplitude of H2. In condition 3 (c), H2–H4 and H4–2 kHz covaried by decreasing the amplitude of H4. Manipulations are represented by dashed lines.

TABLE V. Summary of the five conditions in experiment 2. Arrows indicate an increase (↑) or decrease (↓) for a given component. An unmentioned component means it was held constant in that condition.

Condition	Harmonic varied	Varying components
Condition 1	H1	H1-H2 ↑
Condition 2	H2	H1-H2 ↑, H2-H4 ↓
Condition 3	H4	H2-H4 ↑, H4-2 kHz ↓
Condition 4	2 kHz	H4-2 kHz ↑, 2 kHz-5kHz ↓
Condition 5	5 kHz	2 kHz-5kHz ↑

constant noise-to-signal ratio of −25 dB across all the harmonic-amplitude adjustments.

Sample slope values per stimulus of the first two components can be found in supplementary materials and a summary description of the conditions is shown in Table V. The four components were manipulated in five conditions, each with 18 stimuli, for a total of 90 stimuli. To determine listeners' sensitivity to the different source spectral components and their role in the perception of breathiness, the conditions were designed to provide a range of naturally occurring values for each source spectrum slope. In condition 1, the amplitude of H1 was increased so that the slope of H1–H2 varied from −2 to 15 dB in 1 dB increments. In condition 2, only H1–H2 and H2–H4 varied. The slope of H1–H2 was again varied from −2 to 15 dB (in 1 dB increments), but now by lowering H2 to produce a progressively higher value of H1–H2. Because H2 was manipulated, the slope of H2–H4 increased from about 6 to 23 dB as H1–H2 decreased. The two higher components (H4–2 kHz and 2 kHz–5 kHz) were the same for both conditions 1 and 2. In condition 3, only H2–H4 and H4–2 kHz varied.

The amplitude of H4 was manipulated, such that when H2–H4 increased from 6 to 23 dB, the slope of H4–2 kHz decreased from 27 to 10 dB (H1–H2 remained constant at 8 dB, as shown in Fig. 4). In condition 4, H4–2 kHz and 2 kHz–5 kHz varied. The amplitude of the harmonic nearest 2 kHz was manipulated, such that as H4–2 kHz increased from 10 to 27 dB, the slope of 2 kHz–5 kHz decreased from −2 to 15 dB. In condition 5, only the amplitude of the final harmonic at 5 kHz was manipulated in 1 dB increments. For each stimulus, the slope values for every component fell within the normal range for that component, based on a sample of modeled source spectra for 49 English voices (Kreiman *et al.*, 2011).

Once all these vowels were created, a sample onset /t/ from a naturally produced White Hmong token was spliced onto the beginning of each file. The vowels lacked formant transitions, so splicing of an alveolar burst sounded more natural than a labial to the first and third authors. Real Hmong words with /ta/ occur for all the tones.

2. Participants

The 15 subjects from experiment 1 also participated in experiment 2 after completing experiment 1 and taking a break.

3. Task

Experiment 2, which was also implemented in Praat, consisted of a four-alternative forced-choice identification task, during which participants listened to stimuli varying between /ta/ and /tɔ/ with one of two tones, and then chose which of the four possible words they heard. Both strings can have either the high-falling modal or breathy tone, but we also included two different vowel responses because, when making the stimuli, we noted that some spectral manipulations resulted in a change of vowel quality from more [a]-like to more [ɔ]-like, and these vowel qualities contrast in White Hmong. Therefore, listeners were able to choose between breathy or modal /ta/ or /tɔ/.

As in experiment 1, the possible responses (modal *taj*, *toj* and breathy *tag*, *tog*, all of which are real native words) were displayed on a computer screen in standard Hmong orthography. Listeners could hear a 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. Experiment 2 lasted about 20 min, with 90 stimuli repeated randomly in three blocks, for a total of 270 responses per participant. In total, the two experiments took about 45 min to an hour.

B. Results

1. Source spectrum model parameters and cues to breathiness

To determine which spectral components were used by Hmong listeners to perceive the breathy tone, we fit a logistic mixed-effects regression model to the responses, with tone response (breathy vs modal) as the dependent variable and participant as a random intercept. The model included as fixed effects the four harmonic-amplitude components of the source spectrum model (H1–H2, H2–H4, H4–2 kHz, and 2 kHz–5 kHz) The fixed effects were centered to reduce collinearity between them.

The results of the logistic regression show that H1–H2 and H2–H4 were significant predictors of "breathy-toned" responses (see Table VI). Both of these components have positive estimates, meaning that an increase in either resulted in a significantly higher probability of a "breathy-toned" response. The effects of H1–H2 and H2–H4, both when they varied independently of each other (in conditions 1 and 3) and when they co-varied (condition 2), can be seen in Figs. 5(a) and 5(b). In Fig. 5(a), the proportion of "breathy-toned" responses increased as H1-H2 increased for condition 1 (the black line, where all other components were held constant). However, in condition 2 (the gray line), the same linear increase in H1–H2 did not result in a higher

TABLE VI. Fixed-effects results of logistic model predicting "breathy-toned" responses in experiment 2.

	Coefficient β	SE (β)	Z-score	p-value
Intercept	0.39	0.29	1.35	0.18
H1–H2	0.14	0.01	11.46	<0.0001
H2–H4	0.12	0.01	12.49	<0.0001
H4–2 kHz	0.009	0.01	0.93	0.35
2 kHz–5 kHz	−0.004	0.01	−0.42	0.68

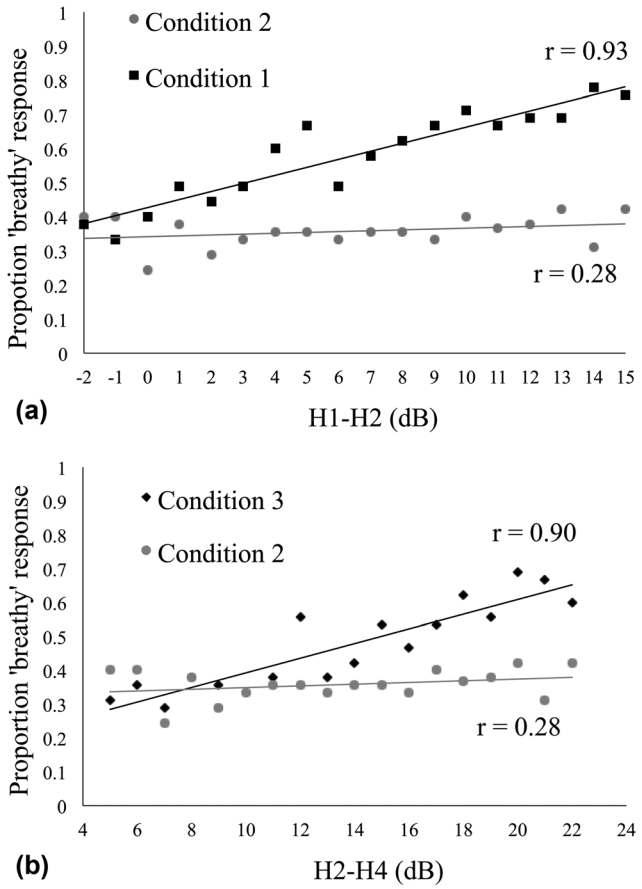


FIG. 5. Proportion of "breathy-toned" responses as a function of (a) H1–H2 in conditions 1 (H1–H2 varying) and 2 (H1–H2 and H2–H4 covarying), (b) H2–H4 in conditions 2 and 3 (H1–H2 held constant, H2–H4 and H4–2 kHz covarying). Best-fit regression lines for proportion of "breathy-toned" responses as a function of changes to (a) H1–H2 and (b) H2–H4 are included, excluding subject-explained variance. H1–H2 manipulations in condition 1 but not in condition 2 are highly correlated with a higher proportion of "breathy-toned" responses (as indicated by the values of Pearson's *r*). H2–H4 manipulations in condition 3 but not in condition 2 are highly correlated with a higher proportion of "breathy-toned" responses. Note that these correlations are not derived from the logistic mixed-effects models.

proportion of "breathy-toned" responses, because in that condition H2–H4 decreased linearly.

However, H2–H4 can independently result in a higher proportion of "breathy-toned" responses, as seen in condition 3 (the black line) in Fig. 5(b), where the effect of H4–2 kHz was minimal. In condition 2 (the gray line), an increase in H2–H4 did not result in more "breathy-toned" responses because as H2–H4 increases, H1–H2 was decreasing. Thus, H1–H2 and H2–H4 may independently trigger a breathy percept in Hmong listeners, but the two components are also in a trading relation. From these results, it might be suspected that the crucial parameter is in fact H1–H4 [as was used by Kingston *et al.* (1997)], subsuming H1–H2 and H2–H4. However, a separate logistic mixed-effects regression model that included H1–H4 as an additional fixed effect showed that both H1–H2 and H2–H4 are still independently significant in predicting "breathy-toned" responses. Thus, for listeners to hear a breathy tone, both an increase in H1–H2 and an increase in H2–H4 are necessary.

2. Source spectrum model parameters and changes in vowel quality

Because participants were also able to choose between two vowels (/a, ɔ/) as well as two phonation types, we next looked at the effects of the source spectrum model's parameters on perceived changes in vowel quality. The filter function was held constant during stimulus creation, so any perceived changes in vowel quality must be due to effects of the source. To determine which spectral components were responsible for perceived changes in vowel quality, we ran a logistic mixed-effects model similar to that used to analyze "breathy-toned" responses, except that in this case the dependent variable was whether listeners chose a word with /ɔ/, as opposed to one with /a/ (regardless of perceived phonation).

The results of the logistic regression show that H1–H2, H2–H4, and H4–2 kHz were significant predictors of /ɔ/ responses (see Table VII). These three components have positive estimates, meaning that increase in the component resulted in a significantly higher probability of obtaining an /ɔ/ response. However, the much higher *Z*-score for H2–H4 suggests that the perceived change in vowel quality is mostly due to that component.

C. Discussion

The results from experiment 2 show that, all else equal, changes in the source spectrum's harmonic amplitudes can affect Hmong listeners' percept of phonemic modal vs breathy voice, which was determined to be contrastive for them in experiment 1. In particular, the results indicate that two of the source spectrum model's components, H1–H2 and H2–H4, are relevant for White Hmong listeners. The results also show that for a consistent breathy percept, the source spectrum's harmonic amplitudes should decrease rapidly from H1 to H4. If there is a sharp drop from H1 to H2, but relatively equal amplitudes of H2, H3, and H4, then listeners will not hear the vowel as breathy. Note that it is not uncommon for voices to have a small H1–H2 and a large H2–H4 (or vice versa). An analysis of 49 English voices indicated that over a third of the sample voices showed such configurations (Kreiman *et al.*, 2011). This appears to be true for Hmong speakers' production of their breathy tone as well. An analysis of 373 Hmong breathy-toned tokens from 36 male and female voices (studied in Esposito, 2012) examined the extent to which H1*–H2* and H2*–H4* are correlated.[5] The measures and the formant correction were implemented in VoiceSauce (Shue *et al.*, 2011). The results show that the correlation between corrected H1*–H2* and

TABLE VII. Fixed-effects results of logistic model predicting "/ɔ/ word" responses in experiment 2.

	Coefficient β	SE (β)	*Z*-score	*p*-value
Intercept	0.97	0.52	1.87	0.06
H1–H2	0.05	0.01	3.80	<0.001
H2–H4	0.15	0.01	13.99	<0.0001
H4–2 kHz	0.04	0.01	3.78	<0.001
2 kHz–5 kHz	0.008	0.01	0.70	0.48

H2*–H4* (averaged over the entire vowel duration) is low, with an r^2 of only 0.02. Therefore, in Hmong a breathy tone can have a small H1–H2 but large H2–H4, and vice versa (regardless of whether the measures derive from the source or audio spectrum). Although this confirms that the spectral configurations in experiment 2 were consistent with naturally occurring breathy tokens, our results do suggest that real breathy tokens with conflicting H1–H2 and H2–H4 slopes would be heard as modal. We discuss this in more detail in Sec. IV B.

IV. GENERAL DISCUSSION

The experiments in this study help clarify the role of phonation in the tonal contrasts of White Hmong, as well as the role of different harmonic amplitudes in the perception of breathy phonation.

A. The role of phonation in the White Hmong tone system

Experiment 1 showed that phonation cues are fundamental for identifying the White Hmong high-falling breathy tone, with F0 modifications having little effect on its perception. That is, breathy voice quality is used to make a categorical distinction between two tones. This is in line with expectations. White Hmong has two tones with very similar high-falling pitch contours, and for these two tones to remain distinct, some difference besides pitch must be made. Modal vs breathy voice qualities provide such a difference, and apparently listeners have come to rely almost exclusively on that difference.

At the same time, however, experiment 1 showed that the role of phonation in the identification of the low-falling creaky tone is comparatively minor, given that an F0 drop and short duration are what listeners relied on in making their judgments. That is, creaky voice quality appears to be at best a secondary cue to what is fundamentally a duration and pitch contrast. Overall rates of identification of the low-falling tone were low in this experiment, suggesting that listeners have rather narrow criteria for the low-falling tone, yet these criteria do not include presence of creaky voice. This result differs from that of Andruski (2006), who found fewer identification errors for naturally occurring creaky-toned stimuli than for modal-toned ones. It is likely that the duration of the creaky-toned stimuli, which is shorter than all other tones, was used as a primary identification cue by listeners in Andruski (2006).

The fact that breathiness is more perceptually important than creakiness might also be due to the slightly greater "burden" of breathiness in the contrast between the two high-falling (j/g-) tones, compared to that of creakiness in the contrast between the two low (s/m-) tones, in the White Hmong lexicon. That is, does the high-falling breathy tone have a higher functional load than the low-falling tone? Based on a dictionary search of White Hmong (Xiong, 2003), 45% of possible minimal pairs between the modal and breathy falling tones exist in the lexicon. A smaller number, 36%, of possible minimal pairs between the low modal and low-falling tones exist. Unfortunately, it is unclear whether this difference in percentage constitutes a meaningfully larger functional load in Hmong for the breathy tone than the creaky tone. We do not know, for example, how frequent these minimal pairs are in the language.

The fact that creaky phonation came out at best as a secondary cue to the low-falling tone might seem surprising given previous evidence that in production this tone's phonation is robustly different from modal (Esposito, 2012; Garellek, 2012). Why would speakers often produce a characteristic phonation if listeners do not expect it or attend to it? And why would White Hmong listeners apparently fail to make use of even the clearly creaky phonation in our stimuli, which should be an informative cue, given that in general listeners make use of any and all relevant information in a speech signal? We propose three possible explanations. First, it is known that rapid dips in F0 can cue the percept of creaky voice quality in Mixtec (Gerfen and Baker, 2005) and of glottal stops in English (Hillenbrand and Houde, 1996), suggesting that some forms of perceived creaky voice can be tied to pitch dynamics alone. The pitch variations in the stimuli in experiment 1 could have produced an integrated percept of creaky voice, even in the absence of physical creak. Nonetheless, this does not explain why listeners ignore creak when it is present.

Another possibility is that F0 is such a salient cue to the low-falling tone, that listeners have come to rely on F0 almost exclusively, despite the relatively small pitch differences between the low-falling creaky and low modal tone contrast.

Alternatively, the key property of the tone might be its low pitch, and creak is simply one means of ensuring a low pitch target. Thus, creak aids speakers in reaching an F0 target, but the target itself is one of pitch. Creaky phonation might also be a consequence of low pitch in general. Likewise, the creaky phonation or a checked-tone with a glottal stop is a means of ensuring that the tone be short in duration. Thus, creakiness likely reinforces both the F0 lowering and the short duration of the low-falling creaky/checked tone but in itself is not distinctive.

The results from experiment 1 also demonstrate differences between creaky and breathy phonation. Although low pitch can be heard as creak, to our knowledge, changes in F0 alone cannot trigger a percept of breathiness. Interestingly, it is common across languages for breathiness to be associated with relatively low or falling tones (Hombert *et al.*, 1979; Gordon and Ladefoged, 2001; Brunelle, 2012), whereas laryngealization (a form of which is creaky voice) can be associated with either lower or higher pitch and tones (Hombert *et al.*, 1979; Gordon and Ladefoged, 2001; Kingston, 2005, and references therein; Brunelle, 2012). Still, creaky voice per se generally means extra-low pitch (Gerratt and Kreiman, 2001). If creaky voice is being used in White Hmong to guarantee an extra-low pitch, then it is functioning more like a pitch setting than like a phonation type that is independent of pitch (Kuang, 2012). On the other hand, breathy voice in White Hmong seems independent of pitch. That is, breathy voice and creaky voice clearly function differently in the White Hmong tone system and that could be because creaky voice is more closely tied to an absolute pitch than breathy voice is.

B. The perceptual importance of harmonic amplitudes

In experiment 2, we found that changes in spectral slopes (as measured by harmonic-amplitude differences) alone are sufficient to change White Hmong listeners' percept from modal voice to breathy voice. These harmonic-amplitude differences were not only sufficient cues, but indeed strong cues, to breathy voice—in these stimuli with only an intermediate level of noise, on its own insufficient to cue breathiness. Of course, we are not claiming that noise is not a cue to breathy voice in White Hmong. It is extremely likely that listeners would attend to noise if it were strong, given the known role of noise in English listeners' perception of breathiness and its interaction with spectral slope (Klatt and Klatt, 1990; Kreiman and Gerratt, 2005; Shrivastav and Sapienza, 2006; Kreiman and Gerratt, 2012) and the measured differences in harmonics-to-noise ratio between the breathy and modal tones in Hmong (Garellek, 2012). However, the current study indicates that when the noise level is only intermediate, spectral slope variation alone can control the rate of breathy responses.

Furthermore, only harmonic-amplitude differences in the low-frequency spectrum, as represented by H1-H2 and H2-H4, mattered to the listeners. Listeners did not use higher frequency modulations to distinguish breathy from modal phonation. This is despite the fact that higher frequency modulations do correlate with breathy vs modal phonations in White Hmong production: Esposito (2012) found that H1*–A2* distinguishes breathy from modal phonation, and Garellek (2010) found that H1*–A3* contributed somewhat to the contrast. H1*–A2* covers frequencies between H1–H4 and part of H4–2k; H1*–A3* will always cover frequencies greater than 2k. Nonetheless, we found that in perception only low frequencies matter. Thus, the spectral slope model is shown to be too detailed for the breathy-modal contrast in White Hmong. This might be surprising, but we can perhaps understand this result by positing that linguistic phonation contrasts, which must be produced by all speakers of a language, may well be less complex than individual voice differences, and so the model needed for linguistic contrasts could be simpler than the model needed for individual voices.

In addition, experiment 2 showed that different harmonic-amplitude parameters (viz. H1–H2 and H2–H4) can independently cue breathiness. This result is consistent with previous work showing that various acoustic measures correlate with production and perception of linguistic breathiness (Esposito, 2010a), and with variation in the slopes of H1–H2 and H2–H4 across individual English voices (Kreiman *et al.*, 2011). What is surprising here is that the perceptual effect of one parameter may undo the effect of the other. The results from experiment 2 suggest that White Hmong breathy-toned vowels with conflicting H1–H2 and H2–H4 slopes would be heard as modal. As shown in Sec. III C., low H1–H2 and high H2–H4 (and vice versa), measured as corrected H1*–H2* and H2*–H4* from audio recordings, do occur in the breathy vowels produced by 36 Hmong speakers, yet it is very unlikely that all those breathy vowels would be consistently misidentified by Hmong listeners.

The source of this paradox is probably the fact that in experiment 2 we held constant other parameters which are likely also important cues to the breathy vs modal contrast in White Hmong. Most importantly, as noted above, spectral noise was held constant, in that the stimuli were created from a modal /a/ with an intermediate level of noise (−25 dB). It is likely that with high levels of noise, tokens with conflicting slopes of H1–H2 and H2–H4 are still heard as breathy. The interaction between H1–H2, H2–H4, and noise should therefore be studied in more detail. The linguistics literature has typically focused on harmonic attributes of breathiness (e.g., Fischer-Jørgensen, 1967; Bickley, 1982; Esposito, 2010a), whereas the voice literature usually focuses on the role of noise (e.g., Hillenbrand *et al.*, 1994; Shrivastav and Sapienza, 2006). The results of this study reinforce that both the harmonic and inharmonic components of the voice source, as well as their interaction, must be important in the perception of phonation, and thus that context is important in the interpretation of acoustic cues to voice quality (Kreiman and Gerratt, 2012).

Spectral slope manipulations, in particular of H2–H4, also resulted in changes in vowel quality identification. Although voice quality modulations may be independent of the filter, researchers have reported vowel quality differences for contrastive phonation types in several languages, possibly due to pharyngeal involvement or larynx movement (Maddieson and Ladefoged, 1985; Denning, 1989; Gordon and Ladefoged, 2001; Edmondson and Esling, 2006; Kuang, 2011; Brunelle, 2012). These authors have shown that in a variety of languages, lax or breathy phonation may co-occur with lower F1 values. Furthermore, perception studies have shown an interdependence of vowel quality and voice quality in listeners' judgments about vowel or voice (Kingston *et al.*, 1997; Lotto *et al.*, 1997; Brunelle, 2012). Such previous findings are consistent with our result that higher spectral tilt caused both more "breathy-toned" and more /ɔ/ (rather than /a/) responses. The higher energy in the lower harmonics might shift listeners' percept of F1 towards the lower end of the frequency scale, even when the filter remains unchanged. That is, vowel height changes in breathy vowels could be perceptually driven, in addition to or instead of physiologically driven. Speakers of languages that have lower F1 values for breathy vowels compared to modal ones might then accentuate this perceived F1 shift by changing the properties of the vocal tract.

In conclusion, we find that breathy phonation is the primary and necessary cue to the high-falling breathy tone in White Hmong, in contrast to creakiness, which (for most listeners) is neither necessary nor sufficient in cueing the low-falling creaky tone. Manipulations of harmonic amplitudes in the source spectrum show that listeners weight a sharp spectral tilt in the lower frequencies as more important than higher-frequency harmonic components for the perception of breathy voice. These results are relevant for determining how many and which spectral parameters are required in a model of the voice source that aims to account for perception of linguistic contrasts as well as cross-voice variability.

ACKNOWLEDGMENTS

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, and Norma Antoñanzas-Barroso for her help with the UCLA voice synthesizer. This work was supported by NSF grants BCS-0720304 and IIS-1018863, and NIH/NIDCD grant DC01797. Supplementary materials may be found on the first author's website.

[1]The harmonic measures are marked with asterisks because they have been corrected for vowel frequencies and bandwidths. For more information, see discussion in Sec. III C.

[2]*Post hoc* within-subject logistic regression analyses reveal that breathy phonation was the sole predictor of "breathy-toned" responses for 12 of the 15 listeners. The remaining 3 listeners' "breathy-toned" responses could not be predicted from the factors included, which suggests either that other factors not studied here accounted for their "breathy-toned" responses, or that the "breathy-toned" responses for these listeners were random.

[3]*Post hoc* within-subject logistic regression analyses reveal that creaky phonation was a significant predictor of "creaky-toned" responses for 2 of the 15 listeners, and had the largest coefficient in the regression model for one listener. The remaining 13 listeners' "creaky-toned" responses were predicted by stimulus duration, mean F0, and F0 slope. This is consistent with the general findings across all listeners, presented in Table IV.

[4]The model's slopes are uncorrected for formants (not marked with asterisks) because they are derived from the *source* spectrum.

[5]The measures here are marked with asterisks to denote that they are corrected for formant frequencies and bandwidths (because they are derived from the audio spectrum). The correction, from Hanson (1997) and Iseli *et al*. (2007), thus approximates the slope amplitudes at the source.

Andruski, J. E. (**2006**). "Tone clarity in mixed pitch/phonation-type tones," J. Phonetics **34**, 388–404.

Andruski, J. E., and Ratliff, M. (**2000**). "Phonation types in production of phonological tone: The case of Green Mong," J. Int. Phonetic Assoc. **30**, 37–61.

Baayen, R. H. (**2008**). *Analyzing Linguistic Data. A Practical Introduction to Statistics* (Cambridge University Press, Cambridge, UK), pp. 1–390.

Belotel-Grenié, A., and Grenié, M. (**1997**). "Types de phonation et tons en chinois standard" ("Phonation types and tones in standard Chinese"), Cah. Ling. **26**, 249–279.

Bickley, C. (**1982**). "Acoustic analysis and perception of breathy vowels," MIT Speech Commun. Working Pap. **1**, 71–81.

Boersma, P., and Weenink, D. (**2011**). "Praat: doing phonetics by computer (version 5.3.02) [computer program]," http://www.praat.org/ (Last viewed November 10, 2011).

Brunelle, M. (**2009**). "Tone perception in Northern and Southern Vietnamese," J. Phonetics **37**, 79–96.

Brunelle, M. (**2012**). "Dialect experience and perceptual integrality in phonological registers: Fundamental frequency, voice quality and the first formant in Cham," J. Acoust. Soc. Am. **131**, 3088–3102.

Brunelle, M., and Finkeldey, J. (**2011**). "Tone perception in Sgaw Karen," in *Proc. ICPhS* 17, 372–375.

Denning, K. (**1989**). "The diachronic development of phonological voice quality, with special reference to Dinka and the other Nilotic languages," Ph.D. thesis, Stanford University, pp. 1–249.

Edmondson, J. A., and Esling, J. H. (**2006**). "The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopic case studies," Phonology **23**, 157–191.

Esposito, C. M. (**2010a**). "The effects of linguistic experience on the perception of phonation," J. Phonetics **38**, 306–316.

Esposito, C. M. (**2010b**). "Variation in contrastive phonation in Santa Ana Del Valle Zapotec," J. Int. Phonetic Assoc. **40**, 181–198.

Esposito, C. M., (**2012**). "An acoustic and electroglottographic study of White Hmong phonation," J. Phonetics **40**, 466–476.

Fischer-Jørgensen, E. (**1967**). "Phonetic analysis of breathy (murmured) vowels in Gujarati," Indian Linguist. **28**, 71–139.

Fulop, S. A., and Golston, C. (**2008**). "Breathy and whispery voice in White Hmong," Proc. Meetings Acoust. **4**, 060006.

Garellek, M. (**2010**). "The acoustics of coarticulated non-modal phonation," UCLA Working Pap. Phonetics **108**, 66–112.

Garellek, M. (**2012**). "The timing and sequencing of coarticulated non-modal phonation in English and White Hmong," J. Phonetics **40**, 152–161.

Garellek, M., and Keating, P. (**2011**). "The acoustic consequences of phonation and tone interactions in Jalapa Mazatec," J. Int. Phonetic. Assoc. **41**, 185–205.

Gerfen, C., and Baker, K. (**2005**). "The production and perception of laryngealized vowels in Coatzospan Mixtec," J. Phonetics **33**, 311–334.

Gerratt, B. R., and Kreiman, J. (**2001**). "Toward a taxonomy of nonmodal phonation," J. Phonetics **29**, 365–381.

Gordon, M., and Ladefoged, P. (**2001**). "Phonation types: A cross-linguistic overview," J. Phonetics **29**, 383–406.

Hanson, H. M. (**1997**). "Glottal characteristics of female speakers: Acoustic correlates," J. Acoust. Soc. Am. **101**, 466–481.

Hillenbrand, J., Cleveland, R. A., and Erickson, R. L. (**1994**). "Acoustic correlates of breathy voice quality," J. Speech Hear. Res. **37**, 769–778.

Hillenbrand, J. M., and Houde, R. A. (**1996**). "Role of F0 and amplitude in the perception of glottal stops," J. Speech Hear. Res. **39**, 1182–1190.

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

Huffman, M. K. (**1987**). "Measures of phonation type in Hmong," J. Acoust. Soc. Am. **81**, 495–504.

Iseli, M., Shue, Y. L., and Alwan, A. (**2007**). "Age, sex, and vowel dependencies of acoustic measures related to the voice source," J. Acoust. Soc. Am. **121**, 2283–2295.

Keating, P., Esposito, C., Garellek, M., Khan, S., and Kuang, J. (**2011**). "Phonation contrasts across languages," in *Proc. ICPhS* 17, 1046–1049.

Kingston, J. (**2005**). "The phonetics of Athabaskan tonogenesis," in *Athabaskan Prosody*, edited by S. Hargus and K. Rice (John Benjamins, Amsterdam), pp. 137–184.

Kingston, J., Macmillan, N. A., Dickey, L. W., Thorburn, R., and Bartels, C. (**1997**). "Integrality in the perception of tongue root position and voice quality in vowels," J. Acoust. Soc. Am. **101**, 1696–1709.

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.

Kreiman, J., Antoñanzas-Barroso, N., and Gerratt, B. R. (**2010**). "Integrated software for analysis and synthesis of voice quality," Behav. Res. Methods **42**, 1030–1041.

Kreiman, J., Garellek, M., and Esposito, C. (**2011**). "Perceptual importance of the voice source spectrum from H2 to 2 kHz," J. Acoust. Soc. Am. **130**, 2570.

Kreiman, J., and Gerratt, B. R. (**2005**). "Perception of aperiodicity in pathological voice," J. Acoust. Soc. Am. **117**, 2201–2211.

Kreiman, J., and Gerratt, B. R. (**2010**). "Perceptual sensitivity to first harmonic amplitude in the voice source," J. Acoust. Soc. Am. **128**, 2085–2089.

Kreiman, J., and Gerratt, B. R. (**2012**). "Perceptual interaction of the harmonic source and noise in voice," J. Acoust. Soc. Am. **131**, 492–500.

Kreiman, J., Gerratt, B., and Antoñanzas-Barroso, N. (**2007**). "Measures of the glottal source spectrum," J. Speech Lang. Hear. Res. **50**, 595–610.

Kuang, J. (**2011**). "Production and perception maps of the multidimensional register contrast in Yi," UCLA Working Pap. Phonetics **109**, 1–30.

Kuang, J. (**2012**). "Registers in tonal contrasts," UCLA Working Pap. Phonetics **110**, 46–64.

Lotto, A. J., Holt, L. L. and Kluender, K. R. (**1997**). "Effect of voice quality on perceived height of English vowels," Phonetica **54**, 76–93.

Maddieson, I., and Ladefoged, P. (**1985**). " 'Tense' and 'lax' in four minority languages of China," J. Phonetics **13**, 433–454.

Michaud, A. (**2004**). "Final consonants and glottalization: New perspectives from Hanoi Vietnamese," Phonetica **61**, 119–146.

Moulines, E., and Charpentier, F. (**1990**). "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," Speech Comm. **9**, 453–467.

Pham, A. H. (**2003**). *Vietnamese Tone: A New Analysis* (Routledge, New York), pp. 1–192.

Ratliff, M. (**1992**). *Meaningful Tone: A Study of Tonal Morphology in Compounds, Form Classes and Expressive Phrases in White Hmong*, Monogr. Ser. Southeast Asia (Northern Illinois University, Center for Southeast Asian Studies, DeKalb, IL), pp. 1–279.

Shrivastav, R., and Sapienza, C. M. (**2006**). "Some difference limens for the perception of breathiness," J. Acoust. Soc. Am. **120**, 416–423.

Shue, Y.-L., Keating, P. A., Vicenik, C., and Yu, K. (**2011**). "VoiceSauce: A program for voice analysis," in *Proc. ICPhS* 17, 1846–1849.
Silverman, D., Blankenship, B., Kirk, P., and Ladefoged, P. (**1995**). "Phonetic structures in Jalapa Mazatec," Anthropol. Ling. **37**, 70–88.
Yu, K. M., and Lam, H. W. (**2011**). "The role of creaky voice in Cantonese tonal perception," in *Proc. ICPhS* 17, 2240–2243.
Xiong, J. (**2003**). *Lus Hmoob Txhais (Hmong-English Dictionary)*; http://www.hmongdictionary.com (Last viewed November 5, 2012).