ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2010
Chicago, IL, USA
May 10-14, 2010

Non-Contrastive Voice Quality Characteristics of Northern Vietnamese Tones

Allison Blodgett, Melissa K. Fox, C. Anton Rytting, Alina Twist

Center for Advanced Study of Language, University of Maryland, College Park, MD, USA
{ablodgett, mfox, crytting, atwist}@casl.umd.edu

Abstract

This study investigated non-contrastive voice qualities in Northern Vietnamese tones with a focus on *huyền* and *hỏi*. The analysis measured the relative amplitude of the first and second harmonics (H1-H2) and of the first harmonic and first formant (H1-A1) in a sample of native speaker speech. While the results are consistent with reports of multiple voice qualities for *huyền* and *hỏi*, *huyền* appeared to be breathy or tense, not breathy or modal. In addition, low falling-rising *hỏi* demonstrated breathy, modal, and tense qualities, while the low falling variant was consistently non-modal.

Index Terms: Vietnamese, lexical tones, voice quality

1. Introduction

Vietnamese orthography reflects six lexical tones: *ngang, huyền, sắc, nặng, hỏi,* and *ngã*. Each tone name contains its corresponding diacritic (or, in the case of *ngang*, no diacritic). Figure 1 illustrates the six tones of Northern Vietnamese [1].

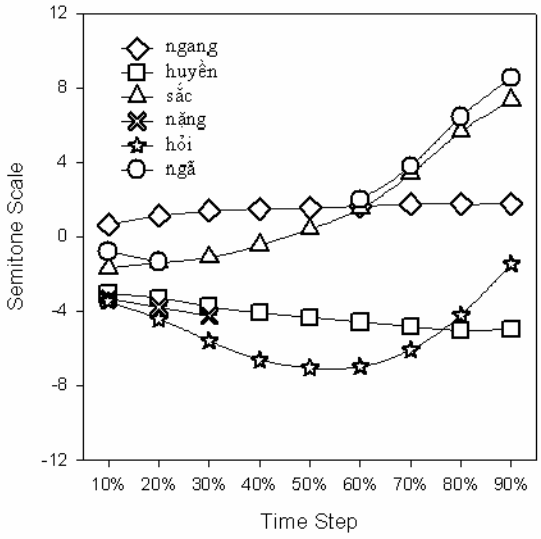

Figure 1: *Time normalized Northern Vietnamese tones from Bauman et al. (2009).*

Northern Vietnamese tones are distinguished by pitch, duration, and voice quality [2,3,4,5,6,7]. Differences between the tones in pitch and voice quality are apparent in Fig. 1. Pitch is represented in semitones on the y-axis. Contrastive voice quality is represented by gaps in tone trajectories for *ngã* and *nặng*. These gaps occur because creaky voice interferes with pitch measurements at these points in the majority of tokens. In Fig. 1 *ngã* shares the same pitch contour as *sắc*, and the creakiness in *ngã* differentiates the two. Similarly, the contours for *nặng* and *huyền* initially align, until creakiness and duration make them distinct.

The creaky gap at the end of Northern *nặng* gives the appearance that this tone is shorter than the others, and duration measurements do indicate that it is significantly shorter than the other tones in Northern Vietnamese [7,8]. However, duration is not actually a feature of Fig 1; pitch measurements have been taken at regular intervals within tone regions in order to normalize variation across tokens.

Creaky voice represents one point on a continuum of voice qualities. The states along the continuum are a function of a number of related properties, such as vocal fold stiffness and degree of glottal constriction [9,10,11]. Whereas the creaky end of the scale involves stiffened vocal folds, narrow glottal constriction, and irregular opening and closing of the vocal folds, the breathy end of the scale involves lax vocal folds, wide glottal opening, and in the extreme, the absence of vocal fold vibration (as in whispered speech). In contrast to these non-modal voice qualities, the central region of the scale – modal voice – maps to regular, periodic opening and closing of the vocal folds. The creakiness in *ngã* and *nặng* can be realized more extremely as one endpoint on the continuum, i.e., as a glottal stop [6].

In the current analysis, a small data set explores reports that Northern Vietnamese *huyền* and *hỏi* exhibit non-modal voice qualities that are not necessarily contrastive or obligatory, but that nonetheless occur with enough regularity in native speaker speech that adult learners might be able to improve their intelligibility if they were to produce those features. More specifically, *huyền* has been reported to be consistently breathy at its end [5] or breathy for some speakers and otherwise modal [4,6,7]. In turn, *hỏi* has been described as tense [4,6], as creaky [2], or as breathy at its midpoint [5]. A further complication for *hỏi* is the fact that it has two known pitch contours [6]. Sometimes it is a low falling-rising tone, and other times it is a low falling tone. There is a strong claim in the literature that speakers consistently produce a non-modal voice quality for low falling *hỏi* such that it is either creaky or strongly breathy [12].

These descriptions establish three main predictions. First, *huyền* will either be breathy or modal for each of four native Northern speakers. Second, *hỏi* will show considerable variation across and within these speakers, consistent with the three voice qualities attributed to it. Third, any low falling variants of *hỏi* will be clearly creaky or breathy.

2. Methodology

2.1. Speakers

Four native Northern speakers (two female, two male) participated. All were originally from Vietnam and had been living in an English-speaking country for 3 to 26 years. They ranged in age from 43 to 73, and all had experience teaching Vietnamese as a foreign language to adults.

2.2. Stimuli

The stimuli in the current analysis represent a subset of targets from a larger study of native speaker and adult learner production of tones and vowels. The full set of targets in this larger study comprised 160 monosyllabic words and covered 11 vowels: *i, ê, e, u, ư, ô, ơ, o, a, â, ă.* The vowels *i, u, ư, ô, ơ,* and *a* appeared with all possible tones for each of three syllable types: open (e.g., *ba, bà, bạ, bá, bả, bã*), stop-final (e.g., *bạt, bát*), and nasal-final (e.g., *bang, bàng, bạnh, báng, bảng, vãn*). The vowels *ê, e,* and *o* appeared only in open syllables. Consistent with Vietnamese phonology, the vowels *â* and *ă* appeared in stop-final and nasal-final syllables only. To the extent possible, targets were matched for initial and final segments within syllable type and within vowel. Consistency in consonant place and/or manner was sacrificed as necessary to ensure that all target stimuli were real words.

Targets for the current analysis consisted of two open syllable tokens of three non-high vowels – *a, ơ, ô* – yielding a total of six tokens per tone per speaker. High vowels were excluded to reduce the likelihood that the position of the first formant would boost the amplitude of the first harmonic [10].

2.3. Procedure

Speakers were recorded in a sound-dampened room using Sound Forge 7.0 (22 kHz, 16 bit, mono), a Yamaha 01V96 digital mixing console with no effects settings, and a Neumann TLM 103 microphone.

Speakers produced three-word sentences in response to individual target words that appeared on a computer screen in red, blue, black, or purple. For example, if the target word *bang* appeared in blue, the speaker said *Từ bang xanh* ("the word *bang* is blue"). Speakers had access to the written color names as they completed eight practice trials and then four lists of words. Lists 1 and 2 each contained 102 targets – including all tokens used in the current analysis – with the vowels *i, u, ư, ô, ơ, a, â,* and *ă*. Lists 3 and 4 each contained 58 targets with the vowels *ê, ê, ơ, o, a, â,* and *ă*. Targets appeared in pseudo-random order such that the vowel, tone, and color of the word always changed from one trial to the next. Three additional targets occurred on Lists 1 and 2. Each was a non-adjacent repetition of an existing target, but in a narrow contrastive context (i.e., in the same color as the immediately preceding word). This added one token each of *i, ư, u, ơ, ô,* and *a*. In addition, *ma* occurred in list-final position on every list. Targets that were paired with *xanh* (blue) and *tím* (purple) on List 1 and List 3 were paired with *đen* (black) and *đỏ* (red), respectively, on List 2 and List 4, and vice versa. Speakers thus produced two repetitions of each target word, but novel utterances each time. In this self-paced task, speakers could repeat any utterance before advancing to the next word. If the experimenters (who are not speakers of Vietnamese) thought they detected an error, e.g., a wrong color term, they directed the speaker to repeat that target utterance at the end of the given list. When speakers did repeat, only the final repetition was analyzed.

2.4. Analysis

Target syllables were annotated within their three-word utterances using Praat [13]. Tone region onsets and offsets were marked based on auditory and visual inspection of each waveform and spectrogram. The beginning of the tone region coincided with vowel onset. The end of the tone region coincided with the end of vowel production, i.e., the word offset. A Praat script automatically assigned nine evenly spaced points between each onset and offset to create time steps within the tone region in 10% increments. Additional scripts created long term average spectra over 40 ms windows centered around the 20%, 50%, and 80% time points. These scripts extracted amplitude measurements at the following spectral peaks – first harmonic (H1), second harmonic (H2), first formant (A1), second formant (A2), and third formant (A3) – and displayed their location, thereby allowing for hand correction as needed.

These acoustic measures are common to voice quality analyses [2,14,15,16] and they reflect the fact that languages and speakers within those languages may have different ways of implementing a given voice quality [9,10,11,17,18]. For example, the relative amplitude of the first and second harmonics is an established correlate of the open portion of the glottal cycle as the vocal folds open and close, but other aspects of vocal fold behavior contribute to voice quality [11].

There are no absolute values that indicate breathy, modal, or creaky voice [11]. Rather, relative amplitudes must be interpreted with respect to particular languages and speakers, as is the case with voice onset time measurements for voiced and voiceless stops [19]. In general, however, breathy voice is expected to have higher values than modal voice, and modal voice is expected to have higher values than creaky voice. Because there is general consensus that *ngang* is modal and *nặng* is creaky in Northern Vietnamese, these tones provide a point of reference for interpreting outcomes. This proves especially useful as the small sample size and the need to consider individual patterns make it difficult to conduct reliable statistical analyses. Ultimately, the analysis focuses on patterns that are consistent or inconsistent with the three main predictions.

The H1-A2 and H1-A3 analyses revealed no patterns of interest and are not presented here.

3. Results

Tables 1 and 2 summarize the mean relative amplitude patterns of H1-H2 and H1-A1, respectively, for each native speaker by tone. Northern *ngã* was excluded because creakiness at its midpoint yielded undefined values. Similar creakiness in *nặng* explains the undefined (Undef) values for the Northern female speakers. Data from the male speakers does include *nặng* as a reference point as Speaker 8 contributed 6, 5, and 3 values at the 20%, 50%, and 80% time steps (out of a maximum of 6 values at each step), and Speaker 9 contributed 6, 3, and 4 values.

3.1. Prediction 1

There is partial support for the prediction that *huyền* would either be breathy or modal for each of the four native Northern speakers. Speaker 15 (female) demonstrated an H1-A1 pattern that is consistent with breathiness for *huyền* relative to *ngang*. That is, the mean values for *huyền* were greater than those for *ngang*, a tone that is generally agreed to be modal. This pattern stands in stark contrast to the remaining Northern speakers, who generally showed lower mean values for *huyền* relative to *ngang*.

Unexpectedly, Speakers 8 and 9 demonstrated H1-A1 and H1-H2 values for *huyền* that approximated those of *nặng*, a tone that is known to be creaky. This suggests that *huyền* is more likely to have tense voice (a quality near the creaky end of the continuum) than modal voice at least for these male speakers. Indeed the *huyền* values for Speaker 9 at the 20%

and 50% time steps approximate those of *sắc*, a tone that has been reported to be tense [6].

Huyền in these cases is probably not creaky given that (1) it would be confusable with *nặng* and (2) creakiness would have interfered with the measurements. Whereas values for *huyền* could be obtained in all but one case, values could only be obtained for 57% of Northern *nặng* cases.

Table 1. *Mean native speaker H1-H2 values.*

		Mean H1-H2 (dB)		
Speaker, Gender	Tone	20%	50%	80%
1, female	***ngang***	**5.38**	**7.31**	**7.59**
	sắc	2.78	5.49	9.58
	hỏi	1.51	6.44	6.10
	huyền	**1.01**	**1.43**	**3.59**
	nặng	-0.73	Undef	Undef
8, male	***ngang***	**10.87**	**12.83**	**15.48**
	sắc	-8.40	7.07	8.02
	hỏi	-9.36	-4.11	-10.97
	huyền	**-13.51**	**-11.53**	**-7.83**
	nặng	-13.42	-9.12	-2.27
9, male	***ngang***	**-0.63**	**2.11**	**2.39**
	sắc	-12.40	-7.00	-1.29
	hỏi	-5.52	2.14	-3.04
	huyền	**-13.44**	**-11.90**	**-9.24**
	nặng	-13.50	-3.77	-2.71
15, female	***ngang***	**4.02**	**3.89**	**14.78**
	sắc	5.21	9.48	12.48
	hỏi	6.17	9.08	6.69
	huyền	**3.09**	**4.55**	**9.84**
	nặng	0.33	Undef	Undef

Table 2. *Mean native speaker H1-A1 values.*

		Mean H1-A1 (dB)		
Speaker, Gender	Tone	20%	50%	80%
1, female	***ngang***	**3.36**	**5.37**	**6.31**
	sắc	2.13	5.07	4.56
	hỏi	4.41	4.54	7.45
	huyền	**6.83**	**4.96**	**5.56**
	nặng	-2.57	Undef	Undef
8, male	***ngang***	**5.01**	**7.00**	**11.33**
	sắc	-6.37	1.16	7.50
	hỏi	-8.95	1.20	-5.15
	huyền	**-9.30**	**-10.67**	**-7.47**
	nặng	-8.58	-4.94	-6.46
9, male	***ngang***	**-1.49**	**-0.82**	**2.63**
	sắc	-15.15	-12.72	-0.82
	hỏi	-5.26	3.15	-0.94
	huyền	**-13.57**	**-13.45**	**-9.70**
	nặng	-18.09	-14.51	-10.82
15, female	***ngang***	**1.28**	**4.99**	**10.55**
	sắc	6.94	7.32	8.76
	hỏi	8.19	11.77	7.20
	huyền	**5.18**	**11.38**	**16.96**
	nặng	-2.53	Undef	Undef

The patterns for Speaker 1 (female) are difficult to interpret. On the one hand, the similarity of the H1-A1 values suggests that *huyền* and *ngang* are modal. On the other hand, lower values for *huyền* relative to *ngang* in the H1-H2 data might suggest that *huyền* is tense. Ideally, a larger data set would test for possible gender-based differences [18].

3.2. Prediction 2

There was support for the prediction that *hỏi* would show considerable variation across speakers. First, Speaker 15 demonstrated a mean value at 50% that was greater for *hỏi* than for *ngang*, consistent with claims of mid-tone breathiness [5]. Speaker 9 demonstrated a similar, but weaker, relationship in the H1-A1 data. Second, Speaker 1 demonstrated midpoint values that are most likely modal given their proximity to *ngang*. Third, Speaker 8 demonstrated H1-H2 and H1-A1 values that may correspond to a tense voice quality given that the values were lower than modal *ngang*, but higher than creaky *nặng*.

Speaker 15 provided additional evidence that *hỏi*'s voice quality can vary even within an utterance: she produced two *hỏi* tokens in a row with distinct voice qualities. The *hỏi* on the middle word contained no audible creakiness and showed continuous pitch tracking. An apparent decrease in amplitude around the tone's midpoint (visible as lightening in the spectrogram) is consistent with a breathy voice quality. The *hỏi* on the third and final word contained audible creakiness, and irregular glottal pulses, which visibly interfered with the pitch tracking.

3.3. Prediction 3

With so few tokens in the analysis, it was impossible to use relative amplitude to examine the prediction that any low falling variants of *hỏi* would clearly be creaky or breathy [12]. Rather, visual inspection of the original set of List 1 and 2 recordings identified low falling variants from each of the native Northern speakers. This process included *hỏi* as it appeared in its target position, i.e., as the second word in a three-word utterance (N=28), and also in its final position as the color term *đỏ* (*red*; N=33). Speaker 15 produced no low falling variants. Speakers 1 and 8 each produced a single candidate, and Speaker 9 produced seven. Six candidates appeared in second position, and three, in third position.

Of the low falling *hỏi* candidates, only Speaker 1's production displayed irregular glottal pulses indicative of creakiness. To assess breathiness in the corresponding tokens from Speakers 8 and 9, the harmonics-to-noise ratio [18] was measured within two low falling tone regions of each utterance: the *huyền* tone in the first word and the *hỏi* tone in the second. Consistent with the claim that these *hỏi* tokens should be breathy [12], *hỏi* tone values (mean 7.62 dB, SD 1.06) were consistently lower than the *huyền* values (mean 14.08 dB, SD 2.14) indicating relatively greater noise in the signal consistent with greater airflow. This difference was statistically significant in a paired samples *t*-test ($t[7]=8.5$, $p<.01$), but is admittedly confounded with utterance position and possibly with pitch height. Additional work is needed to provide a stronger test of this prediction.

4. Discussion

The goal of the current analysis was to explore the non-contrastive voice qualities of *huyền* and *hỏi* using a small set of data culled from a larger experiment focused on tone trajectory and vowel production. Whereas *huyền* was predicted

to be breathy or modal for each of four native Northern speakers, the results of relative amplitude analyses suggested that *huyền* was breathy for one female speaker and tense for at least two others (both male). This is somewhat surprising as there seem to be no prior reports of *huyền* being tense. Only one other study has examined relative amplitudes by tone, using a different methodology and four speakers [2]. Whereas one speaker (female) produced a modal pattern for *huyền*, the other three (1 female, 2 male) produced values that were neither modal nor breathy, perhaps consistent with our conclusion that *huyền* can be tense for some speakers. While the current finding requires replication with a much larger set of dedicated materials, the results do support the conclusion that breathiness is not the norm for Northern *huyền*.

The results also provided support for the prediction that *hỏi* would show considerable variation. Comparisons to the modal *ngang* tone at *hỏi*'s midpoint suggested the use of breathy, modal, and tense voice qualities across speakers. One individual speaker even produced distinct voice qualities within a given utterance.

Despite the variation in voice quality among the *hỏi* tokens with a low falling-rising trajectory (the tone contour represented in Fig. 1), the results provided support for the prediction that the low falling tokens would be consistently non-modal.

The low falling variant of *hỏi* is particularly interesting because it raises the question of how native speakers might distinguish three low falling tones: *huyền*, *nặng*, and *hỏi*. As shown in Fig. 1, *nặng* and *huyền* share a similar trajectory. Two well-established properties likely contribute to make *nặng* distinct: its short duration and its creaky voice. However, if creakiness in the low falling variant of *hỏi* similarly truncates this tone, then either some other attribute(s) must differentiate low falling *hỏi* and *nặng* or the tones are merging. If the tones are not merging, one candidate for differentiation is the relatively lower pitch for *hỏi* (visible in Fig. 1) around its midpoint relative to *nặng* and *huyền*. This relatively lower pitch early in the tone would also potentially distinguish a breathy low falling *hỏi* from a breathy *huyền*.

5. Conclusions

The current study provides native speaker data on the production of non-contrastive voice qualities in Northern Vietnamese tones and highlights promising areas for future investigation. Consistent with reports in the literature, the results demonstrated multiple voice qualities for *huyền* and *hỏi*. Contrary to reports, however, *huyền* appeared to be breathy or tense, not breathy or modal. As expected, variation for *hỏi* occurred not only across speakers, but within speaker and even within utterance. The results also supported claims that the low falling variant of *hỏi* would consistently show creaky or breathy voice.

6. Acknowledgements

We gratefully thank our speakers for their participation, as well as Jessica Bauman, Anita Bowles, Henk Haarmann, Pamela Kling, Sue-Sue Luu, Jessica Shamoo, and Matt Winn for their assistance on the project.

Funding/Support: This material is based upon work supported, in whole or in part, with funding from the United States Government. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the University of Maryland, College Park and/or any agency or entity of the United States Government. Nothing in this report is intended to be and shall not be treated or construed as an endorsement or recommendation by the University of Maryland, United States Government, or the authors of the product, process, or service that is the subject of this report. No one may use any information contained or based on this report in advertisements or promotional materials related to any company product, process, or service or in support of other commercial purposes. The Contracting Officer's Representative for this project is David Cox, Government Technical Director at CASL, (301) 226-8970, dcox@casl.umd.edu.

7. References

[1] Blodgett, A., Bauman, J., Bowles, A., Charters, L., Rytting, A., Shamoo, J., & Winn, M. (2008). A comparison of native speaker and American adult learner Vietnamese lexical tones. *Proceedings of Acoustics '08*, 687-692.

[2] Brunelle, M. (2003). *Coarticulation effects in Northern Vietnamese tones*. Unpublished manuscript, Cornell University.

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

[4] Nguyen, V, & Edmondson, J. (1998). Tones and voice quality in modern northern Vietnamese: Instrumental case studies. *Mon-Khmer Studies*, 28, 1-18.

[5] Pham, A. (2003). *Vietnamese Tone: A new analysis*. New York: Routledge.

[6] Thompson, L. (1965). *A Vietnamese Reference Grammar*. Hawaii: University of Hawaii.

[7] Vu, P. (1981). *The Acoustic and Perceptual Nature of Tone in Vietnamese*. Unpublished doctoral dissertation, Australian National University.

[8] Bauman, J., Blodgett, A., Rytting, C., & Shamoo, J. (2009). *The ups and downs of Vietnamese tones: A description of native speaker and adult learner tone systems for Northern and Southern Vietnamese* (Tech. Rep. No. E.5.3 TTO 2118). College Park, MD: University of Maryland, Center for Advanced Study of Language.

[9] Blankenship, B. (2002). The timing of nonmodal phonation in vowels. *Journal of Phonetics*, 30, 163-191.

[10] Hanson, H. (1997). Glottal characteristics of female speakers: Acoustic correlates. *Journal of the Acoustical Society of America*, 101(1), 466-481.

[11] Keating, P. & Esposito, C. (2006). Linguistic voice quality. *UCLA Working Papers in Phonetics*, 105, 85-91.

[12] Vu Ngoc, T., d'Alessandro, C., & Michaud, A. (2005).Using open quotient for the characterisation of Vietnamese glottalised tones. *Proceedings of Interspeech 2005*, 2885-2888.

[13] Boersma, P. & Weenink, D. (2008). *Praat: Doing Phonetics by Computer*. (Version 4.4.28).

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

[15] Huffman, M. (1987). Measures of phonation type in Hmong. *Journal of the Acoustical Society of America*, 81(2), 495-504.

[16] Watkins, J. (1997). Can phonation types be reliably measured from sound spectra? Some data from Wa and Burmese. *SOAS Working Papers in Linguistics and Phonetics*, 7, 321-339.

[17] DiCanio, C. (2009). The phonetics of register in Takhian Thong Chong. *Journal of the International Phonetic Association*, 39, 162-188.

[18] Wayland, R. & Jongman, A. (2003). Acoustic correlates of breathy and clear vowels: the case of Khmer. *Journal of Phonetics*, 31, 181-201.

[19] Lisker, L. & Abramson, A. (1964). A cross-language study of voicing in initial stops. *Word*, 20, 384-422.