The roles of pitch and phonation in Vietnamese and Mandarin

Irene Vogel[1], *Angeliki Athansopoulou*[2]

[1]University of Delaware, USA

[2]University of Calgary, Canada

ivogel@udel.edu, angeliki@udel.edu

Abstract (200 words)

Pitch and phonation may be used individually or together in languages: some languages do not make systematic use of either property, others use one or the other, and others a combination, where different relationships may hold. We investigate this last option in Mandarin and Vietnamese using substantial, systematically collected corpora, first with auditory and visual (spectrogram) assessment of the presence of Creaky Phonation (CP), then with acoustic and statistical (Binary Logistic Regression) Analyses. We focus on the *sắc* and *ngã* tones in Vietnamese, claimed to contrast in CP not F0, and all four tones of Mandarin, where CP often arises with Tone 3 (dipping), and possibly others. We propose that despite differences in the distribution of F0 and CP, both languages crucially require underlying tonal contrasts, but differ in the source and role of CP. In Mandarin, CP correlates with low F0, as a type of "artifact", resulting in gender differences. In Vietnamese, CP cannot be due to F0, as it appears with high tones; instead, it is an additional "gesture" speakers may introduce along with F0 in producing the *ngã* tone, but need not, as seen in the emergence of two speaker groups based on their use of CP.

Index Terms: tone, phonation, pitch, Mandarin, Vietnamese

1. Introduction

The distribution of pitch and phonation in languages may serve as the basis of a general four-way typology (Table 1). While some languages do not make systematic use of either pitch or phonation (e.g., Spanish), others use one or the other (e.g., Yoruba and Gujurati). Still other languages use both properties (e.g., Jalapa Mazatec, Vietnamese, and Mandarin), but the relationship between the properties may vary in such cases.

Table 1: Typology for Pitch and Phonation contrasts

	No Pitch	**Pitch**
No Phonation	Spanish	Yoruba
Phonation	Gujarati	Jalapa Mazatec, Vietnamese, Mandarin

In languages that use both phonation and pitch, the two properties may vary independently or co-vary [1] (Fig,1). In the former, any combination of the contrastive properties is found: e.g., in Jalapa Mazatec, three phonation categories (creaky, modal, breathy voice) may combine with the three tones (high, mid, low) resulting in nine categories [2]. When the properties co-vary, correlations are observed between pitch and phonation. Often these correlations are due to the phonetic themselves, and to their interaction: e.g., low F0 leads to Creaky Phonation (CP) while high F0 leads to tense phonation, and vice-versa [1]. For example, in Mandarin, the tone with the lowest F0 also often exhibits CP [1]. In Vietnamese, however, the co-variation between pitch and phonation categories is not based on their phonetic properties since CP co-occurs with high F0 values [3].

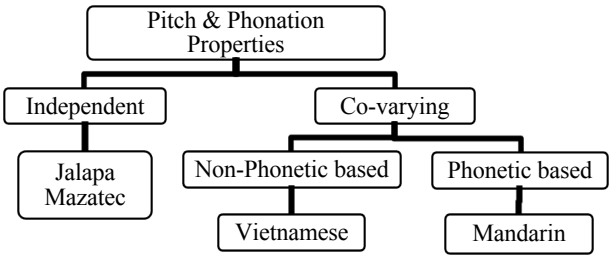

Figure 1. *Typology of languages with pitch and phonation contrasts*

In this study, we investigate the roles of F0 and CP in two systems where they co-vary, Vietnamese and Mandarin.

2. Vietnamese and Mandarin Tones

Northern (Hanoi) Vietnamese has six tones, characterized by different combinations of F0 and phonation properties [3 - 8]. We focus here on the *sắc* and *ngã* tones, which both have rising pitch trajectories, and which have been claimed to be perceptually distinguishable by their phonation difference, as opposed to F0 [5, 6, 7, 9], especially in older speakers. While age may explain some differences in earlier descriptions of the contours (e.g., [10, 11, 12, 13, 4] vs. [3, 14]), in all cases, the *ngã* tone is associated with CP.

The four tones of Mandarin are all characterized as having different pitch patterns; however, it is noted that CP may arise due to low pitch targets in some tones [1]. Most notably, the dipping third tone (T3) often exhibits CP corresponding to the low F0 in the middle. CP has also been reported for the falling fourth tone (T4), again claimed to be due to the low F0 target [2-5]. Both tones are crucially different from the high first tone (T1), which is always modal. The rising second tone (T2), beginning in the mid-range, is also considered to lack CP. Thus, CP in Mandarin does not serve to distinguish between tones, as in Vietnamese, but appears as a type of "artifact" of the low pitch targets, providing enhancement cues for T3 and T4, but not replacing F0 in making crucial tonal contrasts [15, 16, 17].

3. Experimental Design

Given the different claims about pitch and CP in Vietnamese and Mandarin, we expect that acoustic analyses of these properties will also exhibit different distribution patterns. Given the importance of CP in the perception of the Vietnamese *sắc* and *ngã* tones, we expect that it may be crucial in production of the contrast between these tones as well, at the expense of F0.

By contrast, in Mandarin, we expect F0 will be the crucial production property for all four tones, despite any optional addition of CP to T3 and T4. We thus test the hypotheses in Section 3.1 with substantial, systematically controlled corpora from both languages.

3.1. Hypotheses

To investigate the roles of phonation and pitch in Vietnamese and Mandarin, we first verify the presence of CP in the different tones, based on auditory and visual (Praat) assessment.

Hypothesis 1: In Vietnamese the *ngã*, but not sắc, tone has CP.

Hypothesis 2: In Mandarin T3 and T4 have CP; T1 andT2 do not.

We additionally test hypotheses about the acoustic properties, specifically their significance and their relative roles in distinguishing between the various tones in both languages.

Hypothesis 3: In Vietnamese, *sắc* and *ngã* exhibit a distinction in acoustic properties associated with phonation, but not F0.

Hypothesis 4: All Mandarin tones are distinguished by F0.

Hypothesis 5: In Mandarin T3 and T4, but not T1 and T2 exhibit acoustic properties associated with phonation, along with F0.

3.2. Participants

Ten university educated native speakers, aged 18-25 years, were recorded in Hanoi and Beijing, reading short dialogues containing the target items, presented on PowerPoint slides. The participants spoke the standard Hanoi (F=6) or Beijing (F=4) variety of their language. One Vietnamese speaker's data were excluded due to technical issues with the recording.

3.3. Stimuli

As part of a larger cross-linguistic investigation of the acoustic properties of prosody, the stimuli were constructed according to the requirements for all of the languages. Specifically, they are real three-syllable words (compounds in these languages) containing the target vowels /a, i, u/ in the first two syllables. In Vietnamese, we tested 8 items with each vowel and the two rising tones, sắc (modal), ngã (creaky). In Mandarin, we tested 6 items with each vowel and all four tones (Table 2): 96 and 144 per speaker in Vietnamese and Mandarin, respectively.

Table 2: Examples of stimuli with /a/ in syllable 1.

Language	Tone	Examples	
Vietnamese	sắc	*cá voi đực*	'whale' (M)
	ngã	*xã trưởng nữ*	'mayor' (F)
Mandarin	T1	*bā xiān zhuō*	'square table'
	T2	*dá biàn zhuàng*	'reply'
	T3	*dǎ biān gǔ*	'praise'
	T4	*tɑ paɻ tsʰaɻ*	'cabbage'

The stimuli were embedded in short dialogues to prime focus on a word after the target to avoid potential confounds of the acoustic properties of the tones with those of other prosodic phenomena (e.g., focus, boundaries, etc.). (See (1).) The dialogues were adjusted for each tone in Mandarin to minimize tonal coarticulation. Only targets in the answer were examined.

(1) Dialogues for each language. (focused element underlined)

a) Vietnamese: *Ngọc đã nói từ "cá voi đực" hồi sáng à? Không. Ngọc nói từ "cá voi đực" hồi trưa, không phải hồi sáng.*

('Did Ngọc say the word "*cá voi đực*" in the morning? No. Ngọc said the word "*cá voi đực*" in the afternoon, not in the morning.')

b) Mandarin: *Lǎowáng shì biān shuō "bā xiān zhuō" biān pàochá de ma? Búshì Lǎowáng shì biān shuō "bā xiān zhuō" biān xiězì de.*

('Did Laowang say "*bā xiān zhuō*" while making tea? No, Laowang said "*bā xiān zhuō*" while writing.')

3.4. Analyses

Descriptive, acoustic, and statistical analyses were conducted. First, the target vowels were annotated manually in Praat (Boersma & Weenick 2017), providing a *descriptive coding*, indicating the presence of CP based on both auditory and visual information (i.e., spectrograms and waveforms).

Acoustic measurements were made using Voicesauce [18] for the following acoustic properties: duration, mean energy, mean F0, F0 change (ΔF0) (= beginning to end of vowel), F0 change in the first half of the vowel ($\Delta F0_{beg}$) (= beginning to middle of vowel), and F0 change in the second half of the vowel ($\Delta F0_{end}$) (= middle to end of vowel). Three main phonation properties were measured in the middle 1/3 of the vowel: H1-H2, harmonic-to-noise ratio (HNR), and cepstral peak prominence (CPP). The measurements were normalized with z-scores to compensate for vowel and speaker differences.

The acoustic measurements were assessed *statistically* with Binary Logistic Regression Analyses (BLRAs), that first classified (i.e., distinguished between) the various tones in each language using all of the acoustic properties (Overall). Each significant property was then used as the sole classifying property to distinguish the tones, thus assessing its strength as a cue to the tonal distinctions in question (Individual Properties). Only the top two significant predictors are reported below due to space limitations.

4. Results

4.1. Vietnamese

The results from the descriptive analysis unexpectedly revealed two groups of Vietnamese speakers. (See Fig. 2.) One speaker group (Creaky Speakers) used CP in the production of the *ngã* tone (N=6), consistent with previous descriptions [3]; the other group (Modal Speakers) used modal phonation (N=2). The *sắc* tone was modal for both groups. Native speakers who did not participate in our study evaluated the Modal Speaker data and verified that it sounded like normal, naturally produced Northern Vietnamese speech. We thus include the two Modal speakers, but analyze their data separately.

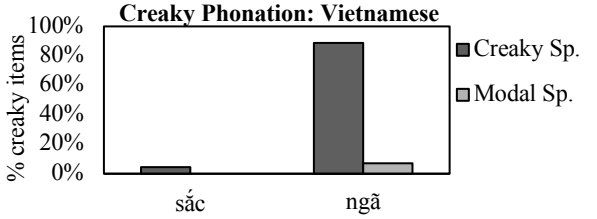

Figure 2. *CP distribution per group in Vietnamese.*

The results of the BLRAs (Table 3) show that both speaker groups successfully distinguish the *sắc* and *ngã* tones (85% - 86% of the time), but differ in which properties are used. The Creaky Speakers use both the phonation property, HNR, and F0 to distinguish the tones, with HNR performing somewhat better than F0 (82% vs. 72%). For the Modal Speakers, only F0 properties were significant (i.e., no phonation properties).

Table 3: Vietnamese BLRAs for Creaky Speakers.

Speaker Group	Overall	Individual properties
Creaky Speakers	85%	***HNR (82%),*** F0 (72%)
Modal Speakers	86%	F0 (88%), ΔF0 (58%)

Fig. 3 shows the HNR patterns: in the Creaky group, HNR, has higher values for the *sắc* than the *ngã* tone, consistent with the descriptive coding for CP. In the Modal group, the two tones do not differ in HNR; both have high HNR values, indicative of modal voice.

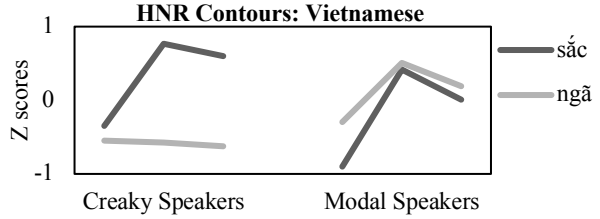

Figure 3. *HNR by speaker group in Vietnamese.*

In Fig.4, the F0 values are also different in the two groups. In the Creaky Speakers, the F0 is only slightly lower for the *sắc* than the *ngã* tone; both follow a similar rising contour. In the Modal Speakers, the F0 is instead considerably lower for the *sắc* than for the *ngã* tone, which also shows a rising contour.

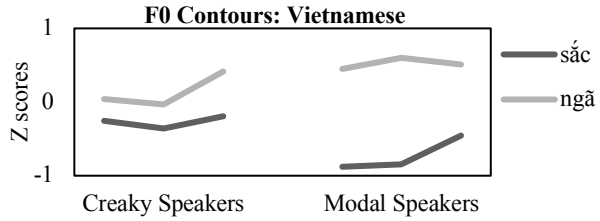

Figure 4. *F0 by speaker group in Vietnamese.*

Thus, while the Creaky Speakers primarily use phonation, but also F0, to distinguish the *sắc* vs. *ngã* tones, the Modal Speakers use only F0, and ensure the distinction with a greater difference between the two tones.

4.2. Mandarin

The descriptive coding of the Mandarin data also revealed two speaker groups, in this case, based on gender (Fig. 5). While the Males (M) use considerable CP with T3, and also some with T4, and even some with T2, the Females (F) use CP only with T3. T1 is modal for both groups of speakers.

The BLRAs (Table 5) show that all of the tonal distinctions are successful (M: 79% - 90%; F73% - 94%). Moreover, for all speakers, either a pitch (F0) property or HNR, the phonation property most commonly appearing with CP, or both, are the top two distinguishing properties for all of the pairs of tones. Consistent with the descriptive coding, HNR is always one of

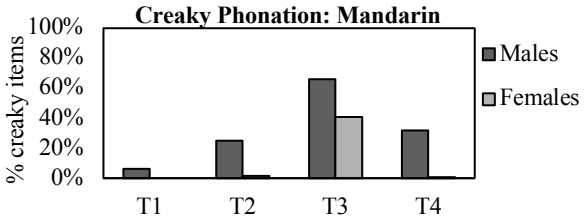

Figure 5. *CP distribution by speaker group in Mandarin.*

the top two properties for all distinctions involving T3 (ii, iv, vi) for F and M. It is also a top property for contrasts involving T4 (iii, v, vi) in M, but not F, paralleling the descriptive results in Fig. 5. The distinction between the two tones described as not having a low target (T1, T2), in fact, does not involve HNR for either group (i).

Table 5: Mandarin BLRAs for Male Speakers.

	Tone Pair	M/F	Overall	Individual properties
i)	T1/T2	M	90%	F0 (79%), $\Delta F0_{end}$ (71%)
		F	94%	F0 (82%), $\Delta F0_{end}$ (68%)
ii)	T1/T3	M	91%	***HNR (80%),*** F0 (76%)
		F	92%	**HNR (82%),** F0 (76%)
iii)	T1/T4	M	84%	$\Delta F0_{beg}$ (72%), ***HNR (72%)***
		F	73%	$\Delta F0_{beg}$ (73%), F0 (65%)
iv)	T2/T3	M	83%	$\Delta F0_{beg}$ (74%), ***HNR (73%)***
		F	84%	$\Delta F0_{beg}$ (82%), ***HNR (72%)***
v)	T2/T4	M	86%	$\Delta F0_{beg}$ (81%), ***HNR (64%)***
		F	91%	$\Delta F0_{beg}$ (75%), F0 (75%)
vi)	T3/T4	M	79%	Enr (68%), ***HNR (66%)***
		F	90%	***HNR (79%),*** CPP (69%)

The HNR patterns (Fig. 6) provide further insight into the distribution of CP.

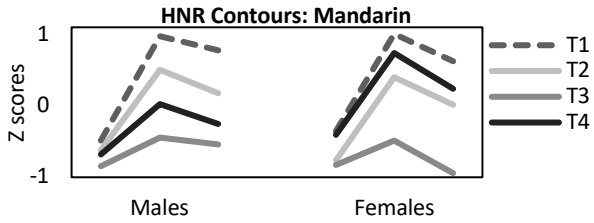

Figure 6. *HNR by speaker group in Mandarin.*

While the HNR values at the middle point (middle 1/3 of vowel) are the lowest for T3 in both M and F, they are high for all of the other tones in F, corresponding to their absence of CP. For M, however, T4 also shows a relatively low HNR, as expected. Additionally, we see that T2 is not as modal as T1, showing some CP (lower HNR), although somewhat less than T3 and T4. This corresponds to the distribution in Fig. 5, but differs from descriptions that indicate that T2 does not exhibit CP.

Despite the role of CP, especially for M, pitch is almost always the main distinguishing property between the tones – except where overwhelmed by the CP of T3 (ii, F vi). This is seen in Fig. 7, where the F0 value of the (middle third of the) vowel is the lowest for T3. F0 is consistently high for T1 for M

and F, and thus the main distinction from the closest tone (T2); the F0 change (i.e., rise) from the middle to the end of the vowel ($\Delta F0_{end}$) also plays a moderate role. The remaining contrasts, however, are most strongly differentiated by the contour, or F0 change, from the beginning to the middle of the vowel ($\Delta F0_{beg}$).

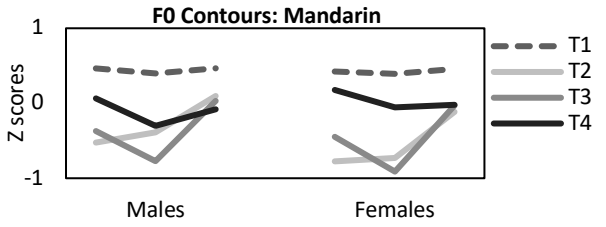

Figure 7. *F0 by speaker group in Mandarin.*

In sum, in Mandarin, CP is a strong cue in the production of T3, as in its perception. The difference in the extent of CP's role in M and F more generally correlates with speakers' low F0, thus appearing in F ***only*** with T3, but also T4 and T2 in M.

5. Discussion

For the most part, our hypotheses have been confirmed with regard to the distribution of CP and its acoustic manifestation, along with different F0 patterns in Vietnamese and Mandarin. For each language, however, we found two groups of speakers who exhibited different CP and F0 patterns, such that we must, in fact, assess the hypotheses differently for each group. Despite internal differences, we see that the relationship between CP and F0 is fundamentally different in the two languages, and thus has broader implications for the analysis of their tonal systems.

With regard to the descriptive analyses, Hypothesis 1 is confirmed for Vietnamese by the Creaky Speakers, who exhibit precisely the expected results: CP with the *ngã* but not the *sắc* tone. The Hypothesis is not, however, confirmed by the Modal Speakers, who do not use CP at all. With regard to Hypothesis 2, in Mandarin, we find that the first part is confirmed, but only in the Males: CP occurs with both T3 and T4, but the Females use CP only with T3. The second part, however, is confirmed only by the F speakers, who do not use CP with either T1 or T2 (but also not with T4); the M speakers do, in fact, use CP with T2, although slightly less than with T4, and T3.

The acoustic analyses parallel the descriptive findings, with HNR (but not other phonation measures) consistently correlating with the observed CP patterns. In Vietnamese, for those speakers who use CP, HNR is high with the *sắc* tone, and low with the *ngã* tone, confirming the first part of Hypothesis 3. For the Modal Speakers, however, HNR is essentially the same (high) for both tones. The second part of the hypothesis is not confirmed by either group, since both do, in fact, consistently produce F0 differences between the two tones, although the difference is considerably greater in the Modal Speakers. Hypothesis 4 is confirmed by both groups in Mandarin, since all of the tones are clearly distinguished by F0 in some way (i.e., overall, change at the beginning or end). Hypothesis 5, however, is only partially confirmed, and must be viewed separately in relation the M and F speakers. Overall, though, where CP is noted in the descriptive coding, we find consistent evidence of the phonation property, low HNR: M, with T3 and T4, as expected, and also T2; but F, only with T3.

Considering the relative roles of CP and F0, the situation seems somewhat simpler in Mandarin. The general pattern is that CP essentially correlates with low F0; since the F speakers use the lowest portion of their pitch range most for T3, this is the only one that exhibits CP. The M speakers not only exhibit CP with T3, but also to a lesser extent with the low target of T4, and the initial low portion of T2. It appears that the difference between the M and F speakers is thus due to the fact that since the former already use a relatively low F0, which could in any case result is some presence of CP, when they produce tones involving low components, the result is the additional presence of CP. The F speakers only reach into their lowest F0 range for T3, and thus produce CP only in this case. For both speaker groups, however, the pattern is the same: CP essentially arises as a physiological "artifact" of low F0. Note that while CP is sometimes claimed to be "allophonic" [1], this cannot, in fact, be the case, since allophones result from regular, rule-governed processes, but CP is not consistently observed, and its presence is not rule-governed, such that in some contexts it always appears, and in others it does not. For all tones, in both M and F, F0 thus provides the main, consistent distinction in production, and it is consistently perceived by listeners.

In Vietnamese, the situation is a bit more complex, since there is a mismatch between what is perceived as the difference between the *sắc* vs. *ngã* tones, and what is produced. While CP the main cue in perception [5], in production, for the Creaky Speakers, it is closely followed by an F0 difference, and for the Modal speakers, the tonal contrast is ***only*** made by F0. It is thus not possible that CP is contrastive, or phonemic, replacing the F0 property of the *ngã* tone. It is also not an "artifact" of low F0, as in Mandarin, since it occurs with high F0 values in both speaker groups. Finally, it also cannot be considered allophonic, since it does not appear regularly in some contexts, and not others, but rather essentially always, or never, occurs with the *ngã* tone, depending on the speaker group – not context. Thus, the tonal system of Vietnamese cannot dispense with F0 as a contrastive property of the *ngã* tone, and indeed, the distinction between this tone and others in the language remains one of F0.

In sum, if we consider the production of CP as a result of a type of articulatory gesture, we can analyze both Mandarin and Vietnamese as tone languages, with tone contrastively represented as abstract F0 patterns. Where the difference thus arises is in the nature of the gestures involved in producing the tones. In Mandarin, no particular "commands" are required to introduce CP, since it comes "for free" with low F0. In Vietnamese, by contrast, CP must be intentionally included along with the F0 commands, allowing some speakers to choose to use it (Creaky Group), and others not to use it (Modal Group). It is possible, moreover, that this approach will turn out to be applicable more generally in other aspects of the Vietnamese tonal system as well, where various phonation properties appear along with F0 properties of different tones.

6. Conclusions

In sum, earlier perceptual studies show that in Vietnamese, CP (not F0) is the crucial contrastive cue for the *sắc* vs. *ngã* tones, but in Mandarin that CP enhances perception of T3, while F0 remains the main cue. Our investigation shows that, despite differences, in production, Vietnamese and Mandarin rely on F0 to make the tonal contrasts. The fundamental difference is that in Mandarin, CP is essentially an "artifact" of the low F0, and thus occurs in more cases with Males than Females, while in Vietnamese, CP occurs with high F0 values, and thus is an additional "gesture" that may, but need not, be used to enhance the *ngã* tone, hence the two speaker groups.

7. References

[1] J. Kuang, Phonation in Tonal Contrasts, Doctoral dissertation, UCLA, 2013.

[2] M. Garellek and P. Keating, "The acoustic consequences of phonation and tone interactions in Jalapa Mazatec," *Journal of the International Phonetic Association,* vol. 41, pp. 185-205, 2011.

[3] M. Brunelle, "Vietnamese (Tiếng Việt)," in *The Handbook of Austroasiatic Languages*, Leiden, Boston, Brill, 2015, pp. 909-953.

[4] M. Brunelle, D. Nguyên and K. Nguyên, "A Laryngographic and Laryngoscopic Study of Northern Vietnamese Tones," *Phonetica,* vol. 67, pp. 147-169, 2010.

[5] M. Brunelle, "Tone perception in Northern and Southern Vietnamese," *Journal of Phonetics,* vol. 37, no. 1, pp. 79-96, 2009.

[6] A. Michaud, "Final consonants and glottalization: New perspectives from Hanoi Vietnamese," *Phonetica,* vol. 61, no. 2-3, pp. 119-146, 2004.

[7] V. Nguyen and J. Edmondson, "Tones and voice quality in modern northern Vietnamese: Instrumental case studies," *Mon-Khmer Studies,* vol. 28, pp. 1-18, 1998.

[8] H. Pham, Vietnamese tone: Tone is not pitch, PhD dissertation, University of Toronto, 2001.

[9] J. Kirby, "Dialect experience in Vietnamese tone perception," *Journal of the Acoustical Society of America,* vol. 127, no. 6, pp. 3749-3757, 2010.

[10] J. Bauman, A. Blodgett, C. Rytting and J. Shamoo, "The ups and downs of Vietnamese tones: A description of native speaker and adult learner tone systems for Northern and Southern Vietnamese," University of Maryland Center for Advanced Study of Language, College Park, MD, 2009.

[11] P. Vu, The acoustic and perceptual nature of tone in Vietnamese, Unpublished doctoral dissertation: Australian National University, 1981.

[12] L. Thompson, A Vietnamese reference grammar, Hawaii: University of Hawaii, 1965.

[13] Đ. T. Dũng, T. T. Hương and G. Boulakia, "Intonation in Vietnamese," in *Intonation Systems: A survey ot twenty languages*, Cambridge Univesity Press, 1998, pp. 398-420.

[14] V. Nguyen and J. Edmondson, "Tones and voice quality in modern Northern Vietnamese: Instrumental case studies," *Mon-Khmer Studies,* vol. 28, pp. 1-18, 1998.

[15] A. Belotel-Grenié and M. Grenié, "Phonation types analysis in Standard Chinese," *ICSLP 3,* 1994.

[16] A. Belotel-Grenié and M. Grenié, "The creaky voice phonation and the organization of Chinese discourse," *TAL,* 2004.

[17] R. Yang, "The role of phonation cues in Mandarin tonal perception," *Journal of Chinese Linguistics,* vol. 43, pp. 453-472, 2015.

[18] Y.-L. Shue, P. Keating, C. Vicenik and K. Yu, "VoiceSauce: A program for voice analysis," *ICPhS XVII,* pp. 1846-1849, 2011.

[19] M. Garellek, P. Keating, C. Esposito and J. Kreiman, "Voice quality and tone identification in White Hmong," *Journal of the Acoustical Society of America,* vol. 133, no. 2, pp. 1078-1089, 2013.

[20] J. Kingston, "Tonogenesis," in *The Blackwell companion to Phonology*, Blackwell, 2011, pp. 2304-2333.

[21] Y. Kang and S. Han, "Tonogenesis in early contemporary Seoul Korean: A longitudinal case study," *Lingua,* vol. 134, pp. 62-74, 2013.

[22] G. Thurgood, "Tonogenesis revisited: Revising the model and the analysis," in *Studies in Tai and Southeast Asian Linguistics*, 2007, pp. 264-291.